

# THEORY OF FUNCTIONAL PRINCIPAL COMPONENT ANALYSIS FOR DISCRETELY OBSERVED DATA

BY HANG ZHOU<sup>1</sup>, DONGYI WEI<sup>2</sup> AND FANG YAO<sup>3\*</sup>

<sup>1</sup>*Department of Statistics, University of California at Davis, [hgzhou@ucdavis.edu](mailto:hgzhou@ucdavis.edu)*

<sup>2</sup>*School of Mathematical Sciences, Peking University, [jnwdyi@pku.edu.cn](mailto:jnwdyi@pku.edu.cn)*

<sup>3</sup>*School of Mathematical Sciences and Center for Statistical Science, Peking University, [fyao@math.pku.edu.cn](mailto:fyao@math.pku.edu.cn)*

Functional data analysis is an important research field in statistics which treats data as random functions drawn from some infinite-dimensional functional space, and functional principal component analysis (FPCA) based on eigen-decomposition plays a central role for data reduction and representation. After nearly three decades of research, there remains a key problem unsolved, namely, the perturbation analysis of covariance operator for diverging number of eigencomponents obtained from noisy and discretely observed data. This is fundamental for studying models and methods based on FPCA, while there has not been substantial progress since [Hall, Müller and Wang \(2006\)](#)'s result for a fixed number of eigenfunction estimates. In this work, we aim to establish a unified theory for this problem, obtaining upper bounds for eigenfunctions with diverging indices in both the  $\mathcal{L}^2$  and supremum norms, and deriving the asymptotic distributions of eigenvalues for a wide range of sampling schemes. Our results provide insight into the phenomenon when the  $\mathcal{L}^2$  bound of eigenfunction estimates with diverging indices is minimax optimal as if the curves are fully observed, and reveal the transition of convergence rates from nonparametric to parametric regimes in connection to sparse or dense sampling. We also develop a double truncation technique to handle the uniform convergence of estimated covariance and eigenfunctions. The technical arguments in this work are useful for handling the perturbation series with noisy and discretely observed functional data and can be applied in models or those involving inverse problems based on FPCA as regularization, such as functional linear regression.

**1. Introduction.** Modern data collection technologies have rapidly evolved, leading to the widespread emergence of functional data that have been extensively studied over the past few decades. Generally, functional data are considered stochastic processes that satisfy certain smoothness conditions or realizations of random elements valued in Hilbert space. These two perspectives highlight the essential natures of functional data, namely, their smoothness and infinite dimensionality, which distinguish them from high-dimensional and vector-valued data. For a comprehensive treatment of functional data, we recommend the monographs by [Ramsay and Silverman \(2006\)](#), [Ferraty and Vieu \(2006\)](#), [Horváth and Kokoszka \(2012\)](#), and [Hsing and Eubank \(2015\)](#), among others.

Although functional data provide information over a continuum, which is often time or spatial location, in reality, data are collected or observed discretely with measurement errors. For instance, we usually use  $n$  to denote the sample size, which is the number of subjects corresponding to random functions, and  $N_i$  to denote the number of observations for the  $i$ th subject. Thanks to the smooth nature of functional data, having a large number of observations per subject is more of a blessing than a curse, in contrast to high-dimensional data

---

\*Fang Yao is the corresponding author.

*MSC2020 subject classifications:* Primary 62R10; secondary 62G20.

*Keywords and phrases:* eigen-decomposition, functional data, perturbation series, phase transition.

(Hall, Müller and Wang, 2006; Zhang and Wang, 2016). There is extensive literature on non-parametric methods that address the smoothness of functional data, including kernel or local polynomial methods (Yao, Müller and Wang, 2005a; Hall, Müller and Wang, 2006; Zhang and Wang, 2016), and various types of spline methods (Rice and Wu, 2001; Yao and Lee, 2006; Paul and Peng, 2009; Cai and Yuan, 2011).

When employing a smoothing method, there are two typical strategies to be considered. If the observed time points per subject are relatively dense, it is recommended to pre-smooth each curve before further analysis, as suggested by Ramsay and Silverman (2006) and Zhang and Chen (2007). However, if the sampling scheme is rather sparse, it is preferred to pool observations together from all subjects (Yao, Müller and Wang, 2005a). The choice of individual versus pooled information affects the convergence rates and phase transitions in estimating population quantities, such as mean and covariance functions. When  $N_i \gtrsim O(n^{5/4})$  and the tuning parameter is optimally chosen per subject, the estimated mean and covariance functions based on the reconstructed curves through pre-smoothing are  $\sqrt{n}$ -consistent, the so-called optimal parametric rate. On the other hand, by borrowing information from all subjects, the pooling method only requires  $N_i \gtrsim O(n^{1/4})$  for mean and covariance estimation to reach optimal (Cai and Yuan, 2010, 2011; Zhang and Wang, 2016), providing theoretical insight into the advantage of the pooling strategy.

However, estimating the mean and covariance functions does not account for the infinite dimensionality of functional data. Due to the non-invertibility of covariance operators for functional random objects, regularization is required in models that involve inverse issues with functional covariates, such as the functional linear model (Yao, Müller and Wang, 2005b; Hall and Horowitz, 2007; Yuan and Cai, 2010), functional generalized linear model (Müller and Stadtmüller, 2005; Dou, Pollard and Zhou, 2012), and functional Cox model (Qu, Wang and Wang, 2016). Truncation of the leading functional principal components (FPC) is a well-developed approach to addressing this inverse issue (Hall and Horowitz, 2007; Dou, Pollard and Zhou, 2012). In order to suppress the model bias, the number of principal components used in truncation should grow slowly with sample size. As a result, the convergence rate of the estimated eigenfunctions with diverging indices becomes a fundamental issue, which is not only important in its own right but also crucial for most models and methods involving functional principal components regularization.

For fully observed functional data, Hall and Horowitz (2007) obtained the optimal convergence rate  $j^2/n$  for the  $j$ th eigenfunction, which served as a cornerstone in establishing the optimal convergence rate in functional linear model (Hall and Horowitz, 2007) and functional generalized linear model (Dou, Pollard and Zhou, 2012). In the discretely observed case, stochastic bounds for a fixed number of eigenfunctions have been obtained by different methods. Using a local linear smoother, Hall, Müller and Wang (2006) showed that the  $\mathcal{L}^2$  rate of a fixed eigenfunction for finite  $N_i$  is  $O_P(n^{-4/5})$ . Under the reproducing kernel Hilbert space framework, Cai and Yuan (2010) claimed that eigenfunctions with fixed indices admit the same convergence rate as the covariance function, which is  $O_P((n/\log n)^{-4/5})$ . It is important to note that, although both results are one-dimensional nonparametric rates (differing at most by a factor of  $(\log n)^{4/5}$ ), the methodologies and techniques used are completely disparate, and a detailed discussion can be found in Section 2. Additionally, Paul and Peng (2009) proposed a reduced rank model and studied its asymptotic properties under a particular setting. In Zhou, Yao and Zhang (2023), the authors studied the convergence rate for the functional linear model and obtained an improved bound for the eigenfunctions with diverging indices. However, this rate will not reach the optimal rate of  $j^2/n$  for any sampling rate  $N_i$ . As explained in Section 2, while some bounds can be obtained for eigenfunctions with diverging indices, attaining an optimal bound presents a substantially greater challenge. Lack of such an optimal bound for eigenfunctions poses considerable challenge in analyzing the

standard and efficient plug-in estimator in functional linear model (Hall and Horowitz, 2007). Consequently, Zhou, Yao and Zhang (2023) resorted to a complex sample-splitting strategy, which results in lower estimation efficiency. To the best of our knowledge, there has been no progress in obtaining the optimal convergence rate of eigenfunctions with diverging indices when the data are discretely observed with noise contamination.

The distinction between estimating a diverging number and a fixed number of eigenfunctions is rooted in the infinite-dimensional nature of functional data. Analyzing eigenfunctions with diverging indices presents challenges due to the decaying eigenvalues. For fully observed data, the cross-sectional sample covariance based on the true functions facilitates the application of perturbation results, as shown in prior work (Hall and Horowitz, 2007; Dou, Pollard and Zhou, 2012). This approach simplifies each term in the perturbation series to the principal component scores. However, when the trajectories are observed at discrete time points, this virtue no longer exists, leading to a summability issue arising from the estimation bias and decaying eigenvalues. This renders existing techniques invalid and remains an unresolved problem; see Section 2 for further elaboration.

This paper addresses this significant yet challenging task of estimating an increasing number of eigenfunctions from discretely observed functional data, and presents a unified theory. The contributions of this paper are at least threefold. First, we establish an  $\mathcal{L}^2$  bound for the eigenfunctions and the asymptotic normality of the eigenvalues with increasing indices, reflecting a transition from nonparametric to parametric regimes and encompassing a wide range from sparse to dense sampling. We show that when  $N_i$  reaches a magnitude of  $n^{1/4+\delta}$ , where  $\delta$  depends on the smoothness parameters of the underlying curves, the convergence rate becomes optimal as if the curves are fully observed. Second, we introduce a novel double truncation method that yields uniform convergence across the time domain, surmounting theoretical barriers in the existing literature. Through this approach, uniform convergence rates for the covariance and eigenfunctions are achieved under mild conditions across various sampling schemes. Notably, this includes the uniform convergence of eigenfunctions with increasing indices, which is new even in scenarios where data are fully observed. Third, we provide a new technical route for addressing the perturbation series of the functional covariance operator, bridging the gap between the “ideal” fully observed scenario and the noisy, discrete “real-world” context. These advanced techniques pave the way for their application in downstream FPCA-related analyses, and the achieved optimal rate of eigenpairs facilitates the extension of existing theoretical results for fully observed functional data to discretely observed case.

The rest of the paper is organized as follows. In Section 2, we give a synopsis of covariance and eigencomponents estimation in functional data. We present the  $\mathcal{L}^2$  convergence of eigenfunctions in Section 3, and discuss the uniform convergence problem of functional data in Section 5. Asymptotic normality of eigenvalues is presented in Section 4. Section 6 provides an illustration of the phase transition phenomenon in eigenfunctions with synthetic data. The proof of Theorem 1 is presented in Appendix A, while Appendix B offers a discussion and insights into the proposed double truncation technique and its application to the uniform convergence of functional data. The proofs of other theorems and ancillary results are collected in the Supplementary Material.

In what follows, we denote by  $A_n = O_P(B_n)$  the relation  $\mathbb{P}(A_n \leq MB_n) \geq 1 - \epsilon$ , and by  $A_n = o_P(B_n)$  the relation  $\mathbb{P}(A_n \leq \epsilon B_n) \rightarrow 0$  as  $n \rightarrow \infty$ , for each  $\epsilon > 0$  and a positive constant  $M$ . A non-random sequence  $a_n$  is said to be  $O(1)$  if it is bounded. For any non-random sequence  $b_n$ , we say  $b_n = O(a_n)$  if  $b_n/a_n = O(1)$ , and  $b_n = o(a_n)$  if  $b_n/a_n \rightarrow 0$ . The notation  $a_n \lesssim b_n$  indicates  $a_n \leq Cb_n$  for sufficiently large  $n$  and positive constant  $C$ , and the relation  $\gtrsim$  is defined similarly. We write  $a_n \asymp b_n$  if  $a_n \lesssim b_n$  and  $b_n \lesssim a_n$ . For  $a \in \mathbb{R}$ ,  $\lfloor a \rfloor$  denotes the largest integer less than or equal to  $a$ . For a function

$f \in \mathcal{L}^2[0, 1]$ , where  $\mathcal{L}^2[0, 1]$  denotes the space of square-integrable functions on  $[0, 1]$ ,  $\|f\|^2$  denotes  $\int_{[0,1]} f(s)^2 ds$ , and  $\|f\|_\infty$  denotes  $\sup_{s \in [0,1]} |f(s)|$ . For a function  $A(s, t) \in \mathcal{L}^2[0, 1]^2$ , define  $\|A\|_{\text{HS}}^2 = \iint_{[0,1]^2} A(s, t)^2 ds dt$  and  $\|A\|_{(j)}^2 = \int_{[0,1]} \{ \int_{[0,1]} A(s, t) \phi_j(s) ds \}^2 dt$ , where  $\{\phi_j\}_{j=1}^\infty$  are the eigenfunctions of interest. We write  $\int pq$  and  $\int Apq$  for  $\int p(u)q(u)du$  and  $\iint A(u, v)p(u)q(v)dudv$  occasionally for brevity.

**2. Eigen-estimation for discretely observed functional data.** Let  $X(t)$  be a square integrable stochastic process on  $[0, 1]$ , and let  $X_i(t)$  be independent and identically distributed (i.i.d.) copies of  $X(t)$ . The mean and covariance functions of  $X(t)$  are denoted by  $\mu(t) = \mathbb{E}\{X(t)\}$  and  $C(s, t) = \mathbb{E}[\{X(s) - \mu(s)\}\{X(t) - \mu(t)\}]$ , respectively. According to Mercer's Theorem (Indritz, 1963),  $C(s, t)$  has the spectral decomposition

$$(1) \quad C(s, t) = \sum_{k=1}^{\infty} \lambda_k \phi_k(s) \phi_k(t),$$

where  $\lambda_1 > \lambda_2 > \dots > 0$  are eigenvalues and  $\{\phi_j\}_{j=1}^\infty$  are the corresponding eigenfunctions, which form a complete orthonormal system on  $\mathcal{L}^2[0, 1]$ . For each  $i$ , the process  $X_i$  admits the so-called Karhunen-Loève expansion

$$(2) \quad X_i(t) = \mu(t) + \sum_{j=1}^{\infty} \xi_{ik} \phi_k(t),$$

where  $\xi_{ik} = \int_0^1 \{X_i(t) - \mu(t)\} \phi_k(t) dt$  are functional principal component scores with zero mean and variances  $\lambda_k$ .

However, in practice, it is only an idealization to have each  $X_i(t)$  for all  $t \in [0, 1]$  to simplify theoretical analysis. Measurements are typically taken at  $N_i$  discrete time points with noise contamination. Specifically, the actual observations for each  $X_i$  are given by

$$(3) \quad \{(t_{ij}, X_{ij}) \mid X_{ij} = X_i(t_{ij}) + \varepsilon_{ij}, j = 1, \dots, N_i\},$$

where  $\varepsilon_{ij}$  are random copies of  $\varepsilon$ , with  $\mathbb{E}(\varepsilon) = 0$  and  $\text{Var}(\varepsilon) = \sigma_X^2$ . We further assume the measurements errors  $\{\varepsilon_{ij}\}_{i,j}$  are independent of  $X_i$ .

Local linear regression is a popular smoothing technique in functional data analysis due to its attractive theoretical properties (Yao, Müller and Wang, 2005a; Hall, Müller and Wang, 2006; Li and Hsing, 2010; Zhang and Wang, 2016). The primary goal of this paper is to develop a unified theory for estimating a larger number of eigenfunctions from discretely observed functional data. To maintain focus and avoid distractions, we assume that the mean function  $\mu(t)$  is known, and set  $\mu(t) = 0$  without loss of generality. The scenario involving an unknown mean function is discussed later in Section 3. We denote by  $\delta_{ijl} = X_{ij} X_{il}$  the raw covariance, and define  $v_i = \{n N_i (N_i - 1)\}^{-1}$ . The local linear estimator of the covariance function is given by  $\hat{C}(s, t) = \hat{\beta}_0$ ,

$$(4) \quad \begin{aligned} (\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2) = \underset{\beta_0, \beta_1, \beta_2}{\text{argmin}} & \sum_{i=1}^n v_i \sum_{1 \leq l_1 \neq l_2 \leq N_i} \{ \delta_{ijl} - \beta_0 - \beta_1 (t_{il_1} - s) - \beta_2 (t_{il_2} - t) \}^2 \\ & \times \frac{1}{h} \text{K} \left( \frac{t_{il_1} - s}{h} \right) \frac{1}{h} \text{K} \left( \frac{t_{il_2} - t}{h} \right), \end{aligned}$$

where  $\text{K}$  is a symmetric, Lipschitz continuous density kernel on  $[-1, 1]$  and  $h$  is the tuning parameter.

The estimated covariance function  $\hat{C}(s, t)$  can be expressed as an empirical version of the spectral decomposition in (1), given by

$$(5) \quad \hat{C}(s, t) = \sum_{k=1}^{\infty} \hat{\lambda}_k \hat{\phi}_k(s) \hat{\phi}_k(t),$$

where  $\hat{\lambda}_k$  and  $\hat{\phi}_k$  are the estimated eigenvalues and eigenfunctions, respectively. These estimates are obtained by solving the eigenequation for  $\hat{C}(s, t)$ :

$$\int \hat{C}(s, t) \hat{\phi}_k(t) dt = \hat{\lambda}_k \hat{\phi}_k(s), \quad \text{with normalization } \int \hat{\phi}_k^2(s) ds = 1.$$

For specificity, we assume  $\langle \hat{\phi}_k, \phi_k \rangle \geq 0$ .

Before delving into the theoretical details, we provide an overview of eigenfunction estimation in functional data analysis. We start with the resolvent series shown in Equation (6) and illustrate its application in statistical analysis,

$$(6) \quad \mathbb{E}(\|\hat{\phi}_j - \phi_j\|^2) \asymp \sum_{k \neq j} \frac{\mathbb{E}[\{\iint (\hat{C} - C) \phi_j \phi_k\}^2]}{(\lambda_k - \lambda_j)^2}.$$

Such expansions can be found in [Bosq \(2000\)](#), [Dou, Pollard and Zhou \(2012\)](#), and [Li and Hsing \(2010\)](#); see Chapter 5 in [Hsing and Eubank \(2015\)](#) for details. Denote by  $\eta_j$  the eigen-gap of  $\lambda_j$ , that is,  $\eta_j := \min_{k \neq j} |\lambda_k - \lambda_j|$ . An basic rough bound for  $\mathbb{E}(\|\hat{\phi}_j - \phi_j\|^2)$  can be derived from Equation (6) and Bessel's inequality,

$$(7) \quad \mathbb{E}(\|\hat{\phi}_j - \phi_j\|^2) \leq \eta_j^{-2} \mathbb{E}(\|\hat{C} - C\|_{\text{HS}}^2).$$

However, this bound is suboptimal for two reasons. First, while  $\eta_j$  is bounded away from 0 for a fixed  $j$ , the bound implies that the eigenfunctions converge at the same rate as the covariance function. This is counterintuitive since integration usually brings extra smoothness ([Cai and Hall, 2006](#)), which typically results in the eigenfunction estimates converging at a faster rate than the two-dimensional kernel smoothing rate of  $\|\hat{C} - C\|_{\text{HS}}^2$ . Second, for  $j$  that diverges with the sample size,  $\eta_j^{-2} \rightarrow \infty$  in the bound cannot be improved. Assuming  $\lambda_j \asymp j^{-a}$ , the rate obtained by (7) is slower than  $j^{2a+2}/n$ , which is known to be suboptimal ([Wahl, 2022](#)). Therefore, to obtain a sharp rate for  $\mathbb{E}(\|\hat{\phi}_j - \phi_j\|^2)$ , one should use the original perturbation series given by (6), rather than its approximation (7).

When each trajectory  $X_i(t)$  is fully observed for all  $t \in [0, 1]$ , the cross-sectional sample covariance  $\hat{C}(s, t) = n^{-1} \sum_{i=1}^n X_i(s) X_i(t)$  is a canonical estimator of  $C(s, t)$ . Then, the numerators in each term of (6) can be reduced to the principal components scores under some mild assumptions, for example,  $\mathbb{E}[\{\iint (\hat{C} - C) \phi_j \phi_k\}^2] \lesssim n^{-1} \lambda_j \lambda_k$  ([Hall and Horowitz, 2007](#); [Dou, Pollard and Zhou, 2012](#)). Subsequently,  $\mathbb{E}(\|\hat{\phi}_j - \phi_j\|^2)$  is bounded by  $(\lambda_j/n) \sum_{k \neq j} \lambda_k / (\lambda_k - \lambda_j)^2$ . With the common assumption of the polynomial decay of eigenvalues, the aforementioned summation is dominated by  $\lambda_j/n \sum_{\lfloor j/2 \rfloor \leq k \leq 2j} \lambda_k / (\lambda_k - \lambda_j)^2$ , which is  $O(j^2/n)$  and optimal in the minimax sense ([Wahl, 2022](#)). See Lemma 7 in [Dou, Pollard and Zhou \(2012\)](#) for a detailed elaboration. This suggests that the convergence rate caused by the inverse issue can be captured by the summation over the set  $\{k \leq 2j\}$ .

However, we would like to emphasize that when it comes to discretely observed functional data, all the existing literature utilizing a bound similar to (7) excludes the case of diverging indices. For instance, the result in [Cai and Yuan \(2010\)](#) is simply a direct application of the bound in (7). Moreover, their one-dimensional rate is inherited from the covariance estimator, which is assumed to be in a tensor product space that is smaller than the space  $\mathcal{L}^2[0, 1]^2$ . On the other hand, the one-dimensional rates obtained by [Hall, Müller and Wang \(2006\)](#) and

Li and Hsing (2010) utilize detailed calculations based on the approximation of the perturbation series in (6). However, these results are based on the assumption that  $\eta_j$  is bounded away from zero, which implies that  $j$  must be a fixed constant. This is inconsistent with the nonparametric nature of functional data models, which aim to approximate or regularize an infinite-dimensional process. Therefore, when dealing with discretely observed functional data, the key to obtaining a sharp bound for estimated eigenfunctions with diverging indices lies in effectively utilizing the perturbation series (6).

The main challenges arise from quantifying the summation in (6) without the fully observed sample covariance. For the pre-smoothing method, the reconstructed  $\hat{X}_i$  achieves a  $\sqrt{n}$  convergence in the  $\mathcal{L}^2$  sense when each  $N_i$  reaches a magnitude of  $n^{5/4}$ , and then the estimated covariance function  $\hat{C}(s, t) = n^{-1} \sum_{i=1}^n \hat{X}_i(s) \hat{X}_i(t)$  has an optimal rate  $\|\hat{C} - C\|_{\text{HS}} = O_P(n^{-1/2})$ . However, this does not guarantee optimal convergence of a diverging number of eigenfunctions. The numerators in each term of (6) are no longer the principal component scores, and the complex form of this infinite summation makes it difficult to quantify when  $|\lambda_k - \lambda_j| \rightarrow 0$ . Similarly, the pooling method also encounters significant challenges in summing all  $\mathbb{E}[\{\iint (\hat{C} - C) \phi_j \phi_k\}^2]$  with respect to  $j$  and  $k$ . Specifically, the convergence rate of  $\|\hat{C} - C\|_{\text{HS}}^2$  should be a two-dimensional kernel smoothing rate with variance  $n^{-1}\{1 + (Nh)^{-2}\}$  (Zhang and Wang, 2016). However, after being integrated twice,  $\mathbb{E}[\{\iint (\hat{C} - C) \phi_j \phi_k\}^2]$  has a degenerated kernel smoothing rate with variance  $n^{-1}$ . According to Bessel's inequality,  $\mathbb{E}(\|\hat{C} - C\|_{\text{HS}}^2)$  can be expressed as  $\sum_{j,k} \mathbb{E}[\{\iint (\hat{C} - C) \phi_j \phi_k\}^2]$ . However, due to estimation bias, one cannot directly sum all  $\mathbb{E}[\{\iint (\hat{C} - C) \phi_j \phi_k\}^2]$  with respect to all  $j, k$ .

**3.  $\mathcal{L}^2$  convergence of eigenfunction estimates.** Based on the issues discussed above, we propose a novel technique for addressing the perturbation series (6) when dealing with discretely observed functional data. To this end, we make the following regularity assumptions.

ASSUMPTION 1. There exists a positive constant  $c_0$  such that  $\mathbb{E}(\xi_j^4) \leq c_0 \lambda_j^2$  for all  $j$ .

ASSUMPTION 2. The second order derivatives of  $C(s, t)$ ,  $\partial C(s, t)/\partial s^2$ ,  $\partial C(s, t)/\partial t^2$  and  $\partial C(s, t)/\partial s \partial t$  are bounded on  $[0, 1]^2$ .

ASSUMPTION 3. The eigenvalues  $\lambda_j$  are decreasing with  $j^{-a} \gtrsim \lambda_j \gtrsim \lambda_{j+1} + j^{-a-1}$  for  $a > 1$  and each  $j \geq 1$ .

ASSUMPTION 4. For each  $j \in \mathbb{N}^+$ , the eigenfunctions  $\phi_j$  satisfies  $\sup_{t \in [0, 1]} |\phi_j(t)| = O(1)$  and

$$\sup_{t \in [0, 1]} |\phi_j^{(k)}(t)| \lesssim j^{c/2} \sup_{t \in [0, 1]} |\phi_j^{(k-1)}(t)|, \quad \text{for } k = 1, 2,$$

where  $c$  is a positive constant.

ASSUMPTION 5.  $\mathbb{E}[\sup_{s \in [0, 1]} |X(t)|^{2\alpha}] < \infty$  and  $\mathbb{E}(\varepsilon^{2\alpha}) < \infty$  for  $\alpha > 1$ .

Assumptions 1 and 2 are widely adopted in the functional data literature related to smoothing (Yao, Müller and Wang, 2005a; Cai and Yuan, 2010; Zhang and Wang, 2016). The decay rate assumption on the eigenvalues provides a natural theoretical framework for justifying and analyzing the properties of functional principal components (Hall and Horowitz, 2007; Dou,

Pollard and Zhou, 2012; Zhou, Yao and Zhang, 2023). The number of eigenfunctions that can be well estimated from exponentially decaying eigenvalues is limited to the order of  $\log n$ , which lacks practical interest. Consequently, we adopt the assumption of polynomial decay in eigenvalues. To achieve quality estimates for a specific eigenfunction, its smoothness should be taken into account. Generally, the frequency of  $\phi_j$  is higher for larger  $j$ , which requires a smaller bandwidth to capture its local variation. Assumption 4 characterizes the frequency increment of a specific eigenfunction via the smoothness of its derivatives. For some commonly used bases, such as the Fourier, Legendre, and wavelet bases,  $c = 2$ . In Hall, Müller and Wang (2006), the authors assumed that  $\max_{1 \leq j \leq r} \max_{s=0,1,2} \sup_{t \in [0,1]} |\phi_j^{(s)}(t)| \leq C$ , which is only achievable for a fixed  $r$ . Therefore, Assumption 4 can be interpreted as a generalization of this assumption. To analyze the convergence of eigenfunctions effectively, a uniform convergence rate of the covariance function is needed to handle the local linear estimator (4). Assumption 5 is the moment assumption required for uniform convergence of covariance function and adopted in Li and Hsing (2010) and Zhang and Wang (2016).

For the observation time points  $\{t_{ij}\}_{i,j}$ , there are two typical types of designs: the random design, in which the design points are random samples from a distribution, and the regular design, where the observation points are predetermined mesh grid. For the random design, the following assumption is commonly adopted (Yao, Müller and Wang, 2005a; Li and Hsing, 2010; Cai and Yuan, 2011; Zhang and Wang, 2016):

**ASSUMPTION 6 (Random design).** The design points  $t_{ij}$ , which are independent of  $X$  and  $\varepsilon$ , are i.i.d. sampled from a distribution on  $[0, 1]$  with a density that is bounded away from zero and infinity.

For the regular design, each sample path is observed on an equally spaced grid  $\{t_j\}_{j=1}^N$ , where  $N_i = \dots = N_n = N$  for all subjects. This longitudinal design is frequently encountered in a various scientific experiments and has been studied in Cai and Yuan (2011). Assumption 7 guarantees a sufficient number of observations within the local window for the kernel smoothing method. For the case where each sample path is observed on the same, but non-equidistant grid  $\{t_j\}_{j=1}^N$ , the results remain applicable by replacing  $N$  in Assumption 7 with  $N^* = 1/\max\{t_j - t_{j-1} : j = 2, \dots, N\}$ . Furthermore, Assumption 8 is needed for the Riemann sum approximation in the fixed regular design.

**ASSUMPTION 7 (Fixed regular design).** The design points  $\{t_j\}_{j=1}^N$  are nonrandom, and there exist constant  $c_2 \geq c_1 > 0$ , such that for any interval  $A, B \in [0, 1]$ ,

- (a)  $c_1 N |A| - 1 \leq \sum_{j=1}^N \mathbb{1}_{\{t_j \in A\}} \leq \max\{c_2 N, 1\}$ ,
- (b)  $c_1 N^2 |A| |B| - 1 \leq \sum_{l_1, l_2}^N \mathbb{1}_{\{t_{l_1} \in A\}} \mathbb{1}_{\{t_{l_2} \in B\}} \leq \max\{c_2 N^2 |A| |B|, 1\}$ ,

where  $|A|$  denotes the length of  $A$ .

**ASSUMPTION 8.**  $\partial C(s, t)/\partial s^2$  and  $\partial C(s, t)/\partial t^2$  are continuously differentiable.

The following theorem is one of our main results. It characterizes the  $\mathcal{L}^2$  convergence of the estimated eigenfunctions with diverging indices for both random design (Assumption 6) and fixed regular design (Assumption 7).

**THEOREM 1.** *Assume Assumptions 1 to 4 hold, further assume Assumption 5 holds with  $\alpha > 3$ .*

(a) For the random design, under Assumption 6, for all  $j \leq m \in \mathbb{N}_+$  satisfies  $m^{2a+2}/n \rightarrow 0$ ,  $m^{2a+2}/(n\bar{N}_2^2 h^2) \rightarrow 0$  and  $h^2 \max\{m^{a+c}, m^{4a} \log n\} \lesssim 1$ ,

$$(8) \quad \|\hat{\phi}_j - \phi_j\|^2 = O_P \left( \frac{j^2}{n} \left\{ 1 + \frac{j^{2a}}{\bar{N}_2^2} \right\} + \frac{j^a}{n\bar{N}_2 h} \left( 1 + \frac{j^a}{\bar{N}_2} \right) + h^4 j^{2c+2} \right),$$

where  $\bar{N}_2 = (n^{-1} \sum_{i=1}^n N_i^{-2})^{-1/2}$ .

(b) For the fixed regular design, under Assumption 7 and 8, for all  $j \leq m \in \mathbb{N}_+$  satisfies  $m^{2a+2}/n \rightarrow 0$ ,  $m^{2a+2}/(nN^2 h^2) \rightarrow 0$ ,  $Nh^a \gtrsim 1$  and  $h^2 \max\{m^{a+c}, m^{4a} \log n\} \lesssim 1$ ,

$$(9) \quad \|\hat{\phi}_j - \phi_j\|^2 = O_P \left( \frac{j^2}{n} + \frac{j^a}{nNh} + h^4 j^{2c+2} \right).$$

To help readers better understand and apply this theory, we include a brief outline and highlight its key elements. Since the local linear estimator is complex, we first extract the dominant component and demonstrate that the remaining terms are dominated by this component. This approach is a standard technique for analyzing kernel estimators. As discussed in Section 2, obtaining the optimal upper bound for eigenfunctions with a diverging indices requires the use of the perturbation series (6). The challenge lies in bounding this summation over the infinite index set  $\{k \neq j\}$  without access to cross-sectional covariance functions in the fully observed case. To address this, one of the key elements is to evaluate the summation in (6) separately over the sets  $S_1 = \{k \leq 2j, k \neq j\}$  and  $S_2 = \{k > 2j\}$ . By Lemma 1 in Appendix A, for each  $k \in S_1$ ,  $\mathbb{E}[\{\iint (\hat{C} - C)\phi_j \phi_k\}^2]$  exhibits a degenerate kernel smoothing rate, with the dominant term scaling as  $j^{-a}k^{-a}/n$ . This allows us to capture the nonparametric rate caused by the inversion issue. For  $S_2$ , instead of bounding each term  $\mathbb{E}[\{\iint (\hat{C} - C)\phi_j \phi_k\}^2]$  individually, we treat the summation  $\sum_{k=2j+1}^{\infty} \mathbb{E}[\{\iint (\hat{C} - C)\phi_j \phi_k\}^2]$  as a unity, which yields a one-dimensional nonparametric rate  $(nNh)^{-1}$ . This approach, together with detailed calculations, results in the proof of Theorem 1. The details of the proof are provided in Appendix A.

The integer  $m$  in Theorem 1 represents the maximum number of eigenfunctions that can be well estimated from the observed data using appropriate tuning parameters. It is important to note that in Theorem 1,  $m$  could diverge to infinity. The upper bound of  $m$  is a function of the sample size  $n$ , the sampling rate  $\bar{N}_2$ , the smoothing parameter  $h$ , and the decaying eigengap  $\eta_j$  or  $a$ . As the frequency of  $\phi_j$  is higher for larger  $j$ , smaller  $h$  is required to capture its local variations. If  $a$  is large, the eigengap  $\eta_j$  diminishes rapidly, posing a greater challenge in distinguishing between adjacent eigenvalues. Note that the assumptions of  $m$  in Theorem 1 could encompass most downstream analyses that involve a functional covariate, such as functional linear regression as discussed in Hall and Horowitz (2007).

Theorem 1 provides a good illustration of both the infinite dimensionality and smoothness nature of functional data. To better understand this result, note that  $j^2/n$  is the optimal rate in the fully observed case. The additional terms on the right-hand side of Equation (8) represent contamination introduced by discrete observations and measurement errors. In particular, the term with  $h^4$  corresponds to the smoothing bias, while the term  $(n\bar{N}_2 h)^{-1}$  reflects the variance typically associated with one-dimensional kernel smoothing. Terms including  $\bar{N}_2^{-1}$  arise from the discrete approximation, and the terms that involve  $j$  with positive powers are due to the decaying eigengaps associated with an increasing number of eigencomponents.

The first part of Theorem 1 provides a unified result for eigenfunction estimates under random design without imposing any restrictions on the sampling rate  $N_i$ . Similar to the phase transitions of mean and covariance functions studied in Cai and Yuan (2011) and Zhang and Wang (2016), Corollary 1 presents a systematic partition that ranges from “sparse” to “dense” schemes for eigenfunction estimation under the random design, which is meaningful for FPCA-based models and methods.

**COROLLARY 1.** *Under Assumptions 1 to 6, and  $m \in \mathbb{N}_+$  satisfies (a) of Theorem 1. For each  $j \leq m$  and let  $h_{opt}(j) = (n\bar{N}_2)^{-1/5} j^{(a-2c-2)/5} (1 + j^a/\bar{N}_2)^{1/5}$ ,*

(a) *If  $\bar{N}_2 \gtrsim j^a$ ,*

$$\|\hat{\phi}_j - \phi_j\|^2 = O_P \left( \frac{j^2}{n} + \frac{j^{(4a+2c+2)/5}}{(n\bar{N}_2)^{4/5}} \right).$$

*In addition, if  $\bar{N}_2 \geq n^{1/4} j^{a+c/2-2}$ ,*

$$\|\hat{\phi}_j - \phi_j\|^2 = O_P \left( \frac{j^2}{n} \right).$$

(b) *If  $\bar{N}_2 = o(j^a)$ ,*

$$\|\hat{\phi}_j - \phi_j\|^2 = O_P \left( \frac{j^{2a+2}}{n\bar{N}_2^2} + \frac{j^{(8a+2c+2)/5}}{(n\bar{N}_2)^{4/5}} \right).$$

Note that  $h_{opt}(j)$  is the optimal bandwidth for estimating a specific eigenfunction  $\phi_j$ . However, in practice, it suffices to estimate the covariance function just once using the optimal bandwidth associated with the largest eigenfunction  $\phi_{m_{\max}}$ . This ensures that both the subspace spanned by the first  $m_{\max}$  eigenfunctions and their corresponding projections are well estimated. When conducting downstream analysis with FPCA, the evaluation of error rates typically involves the term  $\sum_{j=1}^{m_{\max}} j^{-\beta} \|\hat{\phi}_j - \phi_j\|^2$ . Here  $m_{\max}$  denotes the maximum number of principal components used, and  $\beta$  varies across different scenarios, reflecting how the influence of each principal component on the error rate is weighted. For example, in the functional linear model,  $\beta > 0$  represents the rate at which the Fourier coefficients of the regression function decay (Hall and Horowitz, 2007). By the first statement of Theorem 1,

$$\sum_{j=1}^{m_{\max}} j^{-\beta} \|\hat{\phi}_j - \phi_j\|^2 = \frac{m_{\max}^{3-\beta}}{n} \left( 1 + \frac{m_{\max}^{2a}}{\bar{N}_2^2} \right) + \frac{m_{\max}^{a-\beta+1}}{n\bar{N}_2 h} \left( 1 + \frac{m_{\max}^a}{\bar{N}_2} \right) + h^4 m_{\max}^{2c+3-\beta}.$$

It can be found that the optimal bandwidth  $h$  in the above equation aligns with  $h_{opt}(m_{\max})$  for all  $\beta$ . This implies that our theoretical framework can be seamlessly adapted to scenarios where a single bandwidth is employed to attain the optimal convergence rate in the downstream analysis.

For the commonly used bases where  $c = 2$ , the convergence rate for the  $j$ th eigenfunction achieves optimality as if the curves are fully observed when  $\bar{N}_2 > n^{1/4} j^{a-1}$ . Keeping  $j$  fixed, the phase transition occurs at  $n^{1/4}$ , aligning with results in Hall, Müller and Wang (2006) and Cai and Yuan (2010), as well as mean and covariance functions discussed in Zhang and Wang (2016). For  $n$  subjects, the maximum index of the eigenfunction that can be well estimated is less than  $m_{\max} := n^{1/(2a+2)}$ , which directly follows from the assumption  $m_{\max}^{2a+2}/n \rightarrow 0$ . The phase transition in estimating  $\phi_{m_{\max}}$  occurs at  $n^{1/4+(a-1)/(2a+2)}$ . This can be interpreted from two aspects. On one hand, compared to mean and covariance estimation, more observations per subject are required to obtain optimal convergence for eigenfunctions with increasing indices, showing the challenges tied to infinite dimensionality and decaying eigenvalues. On the other hand, the fact that  $n^{1/4+(a-1)/(2a+2)}$  is only slightly larger than  $n^{1/4}$  justifies the merits of the pooling method and supports our intuition. When  $j$  is fixed and  $\bar{N}_2$  is finite, the convergence rate obtained by Corollary 1 is  $(nh)^{-1} + h^4$ , which corresponds to a typical one-dimensional kernel smoothing rate and achieves optimal at  $h \asymp n^{-1/5}$ . This result aligns with those in Hall, Müller and Wang (2006) and is optimal in the minimax sense. When allowing  $j$  to diverge, the known lower bound  $j^2/n$  for fully observed data is attained by applying van Trees' inequality to the special orthogonal group (Wahl, 2022). However,

the argument presented in [Wahl \(2022\)](#) does not directly extend to the discrete case and there are currently no available lower bounds for the eigenfunctions with diverging indices based on discrete observations, which remains an open problem that requires further investigation.

Comparing the results of this work with those in [Zhou, Yao and Zhang \(2023\)](#) is also of interest. The  $\mathcal{L}^2$  convergence rate for the  $j$ th eigenfunction obtained in ([Zhou, Yao and Zhang, 2023](#)) is

$$(10) \quad \frac{j^{a+2}}{n} \left(1 + \frac{1}{Nh}\right) + h^4 j^{2a+2c+2}.$$

It is evident that the results in [Zhou, Yao and Zhang \(2023\)](#) will never reach the optimal rate  $j^2/n$  for any sampling rate  $N$ . In contrast, [Corollary 1](#) provide a systematic partition ranging from “sparse” to “dense” schemes for eigenfunction estimation, and the optimal rate  $j^2/n$  can be achieved when the sampling rate  $N_i$  exceeds the phase transition point. The optimal rate achieved here represents more than just a theoretical improvement over previous findings; it also carries substantial implications for downstream analysis. Further discussion can be found in [Section 7](#).

The following [Corollary](#) presents the phase transition of eigenfunctions for fixed regular design. Note that for fixed regular design, the number of distinct observed time points in an interval of length  $h$  is on the order of  $Nh$ , so  $Nh \gtrsim 1$  is required to ensure there is at least one observation in the bandwidths of kernel smoothing ([Shao, Lin and Yao, 2022](#)). Moreover,  $Nh^a \gtrsim 1$  is required to eliminate the Riemann sum approximation bias. The condition  $Nh^a \gtrsim 1$  is parallel of part (a) in [Corollary 1](#) in the scenario of the random design, which is similar as the mean and covariance estimation where consistency can only be obtained under the dense case for regular design ([Shao, Lin and Yao, 2022](#)).

**COROLLARY 2.** *Under Assumptions 1 to 5, 7, 8 and  $m \in \mathbb{N}_+$  satisfies (b) of [Theorem 1](#). For each  $j \leq m$  and let  $h_{opt}(j) = (nN)^{-1/5} j^{(a-2c-2)/5}$ ,*

$$\|\hat{\phi}_j - \phi_j\|^2 = O_P \left( \frac{j^2}{n} + \frac{j^{(4a+2c+2)/5}}{(nN)^{4/5}} \right).$$

*In addition, if  $N \geq n^{1/4} j^{a+c/2-2}$ ,*

$$\|\hat{\phi}_j - \phi_j\|^2 = O_P \left( \frac{j^2}{n} \right).$$

If the mean function  $\mu(t)$  is unknown, one could use local linear smoother to fit  $\hat{\mu}(t) = \hat{\alpha}_0$  with

$$(\hat{\alpha}_0, \hat{\alpha}_1) = \operatorname{argmin}_{\alpha_0, \alpha_1} \frac{1}{n} \sum_{i=1}^n \frac{1}{N_i} \sum_{j=1}^{N_i} \{X_{ij} - \alpha_0 - \alpha_1(t_{ij} - t)\}^2 \frac{1}{h_\mu} \mathbf{K} \left( \frac{t_{ij} - t}{h_\mu} \right).$$

Then the covariance estimator  $\hat{C}$  is obtained by replacing the raw covariance  $\delta_{ijl}$  in [\(4\)](#) by  $\hat{\delta}_{ijl} = \{X_{ij} - \hat{\mu}(t_{ij})\} \{X_{il} - \hat{\mu}(t_{il})\}$ . The following [corollary](#) presents the convergence rate and phase transition for the case where  $\mu(t)$  is unknown. It should be noted that the Fourier coefficients of  $\mu(t)$  with respect to the eigenfunction  $\phi_j$  generally do not exhibit a decaying trend. Therefore, to eliminate the estimation error caused by the mean estimation, an additional lower bound on  $N$  is necessary. This lower bound, denoted as  $N \gtrsim j^a$ , aligns with the partition described in [Corollary 1](#) for the random design case.

**COROLLARY 3.** *Suppose that Assumptions 1 to 5 hold. Under either of Assumption 6 or 7 and 8, for all  $m \in \mathbb{N}_+$  satisfies  $m^{2a+2}/n$ ,  $Nh^a \gtrsim 1$  and  $h^2 \max\{m^{a+c}, m^{4a} \log n\} \lesssim 1$  and each  $j \leq m$ ,*

$$\|\hat{\phi}_j - \phi_j\|^2 = O_P \left( \frac{j^2}{n} + \frac{j^{(4a+2c+2)/5}}{(n\bar{N}_2)^{4/5}} \right).$$

*In addition, if  $\bar{N}_2 \geq n^{1/4} j^{a+c/2-2}$ ,*

$$\|\hat{\phi}_j - \phi_j\|^2 = O_P \left( \frac{j^2}{n} \right).$$

**4. Asymptotic normality of eigenvalue estimates.** The distribution of eigenvalues plays a crucial role in statistical learning and is of significant interest in the high-dimensional setting. Random matrix theory provides a systematic tool for deriving the distribution of eigenvalues of a squared matrix (Anderson, Guionnet and Zeitouni, 2010; Pastur and Shcherbina, 2011), and has been successfully applied in various statistical problems, such as signal detection (Nadler, Penna and Garello, 2011; Onatski, 2009; Bianchi et al., 2011), spiked covariance models (Johnstone, 2001; Paul, 2007; El Karoui, 2007; Ding and Yang, 2021; Bao et al., 2022), and hypothesis testing (Bai et al., 2009; Chen and Qin, 2010; Zheng, 2012). For a comprehensive treatment of random matrix theory in statistics, we recommend the monograph by Bai and Silverstein (2010) and the review paper by Paul and Aue (2014).

Despite the success of random matrix theory in high-dimensional statistics, its application to functional data analysis is not straightforward due to the different structures of functional spaces. If the observations are taken at the same time points  $\{t_j\}_{j=1}^N$  for all  $i$ , one can obtain an estimator for  $\Sigma_T = \text{Cov}(\tilde{X}_i, \tilde{X}_i) + \sigma_X^2 I_N$ , where  $\tilde{X}_i = (X_i(t_1), \dots, X_i(t_N))^\top$ . Note that  $\tilde{X}_i$  can be regarded as a random vector; however the adjacent elements in  $\tilde{X}_i$  are highly correlated as  $N$  increases due to the smooth nature of functional data. This correlation violates the independence assumption required in most random matrix theory settings.

In the context of functional data, variables of interest become the eigenvalues of the covariance operator. However, the rough bound obtained by Weyl's inequality,  $|\hat{\lambda}_k - \lambda_k| \leq \|\Delta\|_{\text{HS}}$ , is suboptimal from two respects. First,  $|\hat{\lambda}_k - \lambda_k|$  should have a degenerated kernel smoothing rate with variance  $n^{-1/2}$ , whereas  $\|\Delta\|_{\text{HS}}$  has a two-dimensional kernel smoothing rate with variance  $(n\bar{N}_2^2 h^2)^{-1/2}$ . Second, due to the infinite dimensionality of functional data, the eigenvalues  $\{\lambda_k\}_k$  tend to zero as  $k \rightarrow \infty$ , so a general bound for all eigenvalues provides little information for those with larger indices. Although expansions and asymptotic normality have been studied for a fixed number of eigenvalues, as well as for those with diverging indices for fully observed functional data, the study of eigenvalues with a diverging index under the discrete sampling scheme remains limited.

In light of the aforementioned issues, we employ the perturbation technique outlined in previous sections to establish the asymptotic normality of eigenvalues with diverging indices, which holds broad implications for inference problems in functional data analysis. Before presenting our results, we introduce the following assumption, which is standard in FPCA for establishing asymptotic normality.

**ASSUMPTION 9.**  $\mathbb{E}(\|X\|^6) < \infty$  and  $\mathbb{E}(\epsilon^6) < \infty$ . For any sequence  $j_1, \dots, j_4 \in \mathbb{N}_+$ ,  $\mathbb{E}(\xi_{j_1} \xi_{j_2} \xi_{j_3} \xi_{j_4}) = 0$  unless each index  $j_k$  is repeated.

**THEOREM 2.** *Under Assumptions 1 to 4, 6 and 9, for all  $j \leq m \in \mathbb{N}_+$  satisfy  $m = o(n^{1/(2a+4)})$ ,  $hm^{2a} \log n \lesssim 1$  and  $h^4 m^{2a+2c} \lesssim 1$*

$$\Sigma_n^{-\frac{1}{2}} \left( \frac{\hat{\lambda}_j - \lambda_j}{\lambda_j} - \mathbf{K}_2 h^2 \int \phi_j^{(2)}(u) \phi_j(u) du + o(h^2) \right) \xrightarrow{d} \mathcal{N}(0, 1),$$

where

$$\begin{aligned} \Sigma_n &= \frac{4!P_0(N)}{n} \frac{\mathbb{E}(\xi_j^4) - \lambda_j^2}{\lambda_j^2} + 4! \frac{P_1(N)}{n} \frac{\int \{C(u, u) + \sigma_X^2\} \frac{\phi_j^2(u)}{f(u)} du}{\lambda_j} \\ &+ 4 \frac{P_2(N)}{n} \frac{1}{\lambda_j^2} \left( \left[ \int \{C(u, u) + \sigma_X^2\} \frac{\phi_j^2(u)}{f(u)} du \right]^2 - \iint C(u, v)^2 \frac{\phi_j^2(u)\phi_j^2(v)}{f(u)f(v)} dudv \right) \end{aligned}$$

with  $K_2 = \int u^2 K(u) du$  and

$$P_0(N) = \frac{1}{n} \sum_{i=1}^n \frac{(N_i - 2)(N_i - 3)}{N_i(N_i - 1)}, \quad P_1(N) = \frac{1}{n} \sum_{i=1}^n \frac{(N_i - 2)}{N_i(N_i - 1)}, \quad P_2(N) = \frac{1}{n} \sum_{i=1}^n \frac{1}{N_i(N_i - 1)}.$$

Since  $\lambda_j$  approaches zero as  $j$  approaches infinity, we need to regularize the eigenvalues so that they can be compared on the same scale of variability. For a fixed  $j$ ,  $(\hat{\lambda}_j - \lambda_j)/\lambda_j$  is  $\sqrt{n}$ -consistent when  $h$  is small. For diverging  $j$ , Corollary 4 presents three different types of asymptotic normalities, depending on the value of  $P_2(N)$ .

**COROLLARY 4.** *Under the assumptions of Theorem 2,*

(a) *If  $P_2(N)/\lambda_j^2 \rightarrow 0$ ,  $\sqrt{n}h^2 \int \phi_j^{(2)}(u)\phi_j(u)du \rightarrow 0$ ,*

$$\sqrt{n} \left( \frac{\hat{\lambda}_j - \lambda_j}{\lambda_j} \right) \xrightarrow{d} \mathcal{N} \left( 0, 4! \frac{\mathbb{E}(\xi_j^4) - \lambda_j^2}{\lambda_j^2} \right).$$

(b) *If  $P_2(N)/\lambda_j^2 \rightarrow C_3$  for a positive  $C_3$ ,*

$$\begin{aligned} \sqrt{n} \left( \frac{\hat{\lambda}_j - \lambda_j}{\lambda_j} - K_2 h^2 \int \phi_j^{(2)}(u)\phi_j(u)du \right) &\xrightarrow{d} \mathcal{N} \left( 0, 4! \frac{\mathbb{E}(\xi_j^4) - \lambda_j^2}{\lambda_j^2} \right. \\ &+ 4! \sqrt{C_3} \int \{C(u, u) + \sigma_X^2\} \frac{\phi_j^2(u)}{f(u)} du \\ &\left. + 4C_3 \left\{ \left[ \int \{C(u, u) + \sigma_X^2\} \frac{\phi_j^2(u)}{f(u)} du \right]^2 - \iint C(u, v)^2 \frac{\phi_j^2(u)\phi_j^2(v)}{f(u)f(v)} dudv \right\} \right) \end{aligned}$$

(c) *If  $P_2(N)/\lambda_j^2 \rightarrow \infty$ ,*

$$\begin{aligned} \sqrt{\frac{n\lambda_j^2}{P_2(N)}} \left( \frac{\hat{\lambda}_j - \lambda_j}{\lambda_j} - K_2 h^2 \int \phi_j^{(2)}(u)\phi_j(u)du \right) \\ \xrightarrow{d} \mathcal{N} \left( 0, 4 \left\{ \left[ \int \{C(u, u) + \sigma_X^2\} \frac{\phi_j^2(u)}{f(u)} du \right]^2 - \iint C(u, v)^2 \frac{\phi_j^2(u)\phi_j^2(v)}{f(u)f(v)} dudv \right\} \right). \end{aligned}$$

Compared to the mean and covariance estimators, which are associated with one-dimensional and two-dimensional kernel smoothing rates respectively, the estimator of eigenvalues exhibits a degenerate rate after being integrated twice. This implies that there is no trade-off between bias and variance in the bandwidth  $h$ , and the estimation bias can be considered negligible for small values of  $h$ . In this scenario, the phase transition is entirely

determined by the relationship between  $P_2(N)$  and  $\lambda_j$ . Specifically, in the dense and ultra-dense cases where  $P_2(N)/\lambda_j^2 < \infty$ , the variance terms resulting from discrete approximation are dominated by  $1/\sqrt{n}$ , corresponding to cases (a) and (b) in Corollary 4. On the other hand, when each  $N_i$  is relatively small, the estimation variance  $\sqrt{n\lambda_j^2/P_2(N)}$  arising from discrete observations dominates, as outlined in case (c) of Corollary 4. The phase transition point  $j^a$  is the same as the case of eigenfunctions outlined in Corollary 1 and 7, and for larger values of  $j$ , more observations are needed due to the vanishing eigengap.

Using similar arguments, we establish the  $\mathcal{L}^2$  convergence rate for eigenvalues, which is useful for analyzing models involving functional principal component analysis, such as the plug-in estimator in Hall and Horowitz (2007).

**COROLLARY 5.** *Under Assumptions 1 to 4 and 6, for all  $j \leq m \in \mathbb{N}_+$  satisfy  $m = o(n^{1/(2a+4)})$ ,  $hm^{2a} \log n \lesssim 1$  and  $h^4 m^{2a+2c} \lesssim 1$ ,*

$$(\hat{\lambda}_j - \lambda_j)^2 = O_P \left( \frac{j^{-2a}}{n} \{P_0(N) + j^{2a} P_2(N)\} + h^4 j^{2c-2a} \right).$$

**5. Uniform convergence of covariance and eigenfunction estimates.** Classical non-parametric regression with independent observations has yielded numerous results for the uniform convergence of kernel-type estimators (Bickel and Rosenblatt, 1973; Hardle, Janssen and Serfling, 1988; Härdle, 1989). For functional data with in-curve dependence, Yao, Müller and Wang (2005a) obtained a uniform bound for mean and covariance pooling estimates. More recently, Li and Hsing (2010) and Zhang and Wang (2016) have studied the strong uniform convergence of these estimators, showing that these rates depend on both the sample size and the number of observations per subject. However, uniform results for estimated eigenfunctions with diverging indices have not been obtained, even in the fully observed case.

Even for the covariance estimates, there remains a theoretical challenge in achieving uniform convergence for the covariance function under the dense/ultra-dense schemes. Specifically, to obtain uniform convergence with in-curve dependence in functional data analysis, a common approach involves employing the Bernstein inequality to obtain a uniform bound over a finite grid of the observation domain. This grid becomes increasingly dense as the sample size grows. The goal is then to demonstrate that the bound over the finite grid and the bound over  $[0, 1]$  are asymptotically equivalent. This technique has been studied by Li and Hsing (2010) and Zhang and Wang (2016) to achieve uniform convergence for mean and covariance estimators based on local linear smoothers. However, due to the unboundedness natural of functional data, truncation on the observed data is necessary to apply the Bernstein inequality; that is,  $X_{ij}X_{il}\mathbb{1}_{\{X_{ij} \leq A_n\}}$ , where  $A_n \rightarrow \infty$  as  $n \rightarrow \infty$ . The choice of the truncation sequence  $A_n$  should balance the trade-off between the estimation variance appearing in the Bernstein inequality and the bias resulting from truncation. Once the optimal  $A_n$  is chosen, it is essential to impose additional moment conditions on both  $X_i$  and  $\varepsilon_i$  to ensure that the truncation bias is negligible. For covariance estimation, the current state-of-the-art results (Zhang and Wang, 2016) require assuming that the  $6/(1-\tau)$ th moments of  $X_i$  and  $\varepsilon_i$  are finite, where  $\tau$  is the sampling rate with  $\bar{N}_2 \asymp n^{\tau/4}$  for  $\tau \in [0, 1)$ . However, when  $\bar{N}_2$  is close to  $n^{1/4}$ , the value of  $6/(1-\tau)$  tends to infinity. Moreover, in the dense and ultra-dense sampling schemes where  $\bar{N}_2 \gtrsim n^{1/4}$ , the discrepancy between the truncated and the original estimators becomes dominant, preventing the achievement of optimal convergence rates with the current methods.

We first resolve this issue for covariance estimates and introduce an additional truncation step on the summation of random quantities after initially truncating  $X_{ij}X_{il}$  on  $A_n$ . This approach leverages Bernstein's inequality twice to achieve a sharp bound while allowing for

a larger  $A_n$ , thereby reducing the bias introduced by the first truncation. This novel double truncation technique provides a unified framework applicable to all sampling schemes and addresses key limitations in the existing literature. A detailed discussion of the uniform convergence problem for functional data and the double truncation technique is provided in Appendix B.

**THEOREM 3.** *Under the Assumptions 1, 5 and 6,*

(a) *For  $p, q = 0, 1$ , denote*

$$R_{pq}(s, t) = \sum_{i=1}^n v_i \sum_{l_1 \neq l_2}^{N_i} \frac{1}{h^2} \mathbf{K} \left( \frac{t_{il_1} - s}{h} \right) \mathbf{K} \left( \frac{t_{il_2} - t}{h} \right) \left( \frac{t_{il_1} - s}{h} \right)^p \left( \frac{t_{il_2} - t}{h} \right)^q \delta_{il_1 l_2}.$$

*Then*

$$(11) \quad \begin{aligned} & \sup_{s, t \in [0, 1]} |R_{pq}(s, t) - \mathbb{E}R_{pq}(s, t)| \\ & = O \left( \sqrt{\frac{\ln n}{n}} \left( 1 + \frac{1}{\bar{N}_2 h} \right) + \left| \frac{\ln n}{n} \right|^{1 - \frac{1}{\alpha}} \left| 1 + \frac{\ln n}{\bar{N}_2 h} \right|^{2 - \frac{2}{\alpha}} h^{-\frac{2}{\alpha}} \right) a.s. \end{aligned}$$

(b) *In addition, if Assumption 2 holds and  $\alpha > 3$ ,*

(1). *When  $\bar{N}_2 / (n / \log n)^{1/4} \rightarrow 0$  and  $h \asymp (n \bar{N}_2^2 / \log n)^{-1/6}$ ,*

$$\sup_{s, t \in [0, 1]} |\hat{C}(s, t) - C(s, t)| \lesssim \left( \frac{\log n}{n \bar{N}_2^2} \right)^{\frac{1}{3}} a.s.$$

(2). *When  $\bar{N}_2 / (n / \log n)^{1/4} \rightarrow C_1$  for a positive constant  $C_1$  and  $h \asymp (n / \log n)^{1/4}$*

$$\sup_{s, t \in [0, 1]} |\hat{C}(s, t) - C(s, t)| \lesssim \sqrt{\frac{\log n}{n}} a.s.$$

(3). *When  $\bar{N}_2 / (n / \log n)^{1/4} \rightarrow \infty$ ,  $h = o(n / \log n)^{1/4}$  and  $\bar{N}_2 h \rightarrow \infty$ ,*

$$\sup_{s, t \in [0, 1]} |\hat{C}(s, t) - C(s, t)| \lesssim \sqrt{\frac{\log n}{n}} a.s.$$

The first statement of Theorem 3 establishes the uniform convergence rate of the variance term of  $\hat{C} - C$  for all sampling rates  $\bar{N}_2$ . It is worth comparing our results with those in Li and Hsing (2010) and Zhang and Wang (2016). The bias term caused by double truncation, which appears in the second term on the right-hand side of (11), is smaller than those obtained in Li and Hsing (2010) and Zhang and Wang (2016). As a result, the second statement of Theorem 3 shows that the truncation bias is dominated by the main term if  $\alpha > 3$ , a condition that is much milder compared to the moment assumptions in Li and Hsing (2010) and Zhang and Wang (2016), thereby establishing the uniform convergence rate for both sparse and dense functional data. Moreover, when  $1 < \alpha \leq 3$ , which corresponds to the case where  $X(t)$  or  $\varepsilon$  do not possess a sixth order finite moment, the additional term caused by truncation becomes dominant. In summary, by introducing the double truncation technique, we resolve the aforementioned issues present in the original proofs of Li and Hsing (2010) and Zhang and Wang (2016), achieve the uniform convergence rate for the covariance estimator across all sampling schemes, and show that the optimal rates for dense functional data can be obtained as a special case. Having set the stage, we arrive at the following result that gives the uniform convergence for estimated eigenfunctions with diverging indices.

**THEOREM 4.** *Under Assumptions 1 to 5, for all  $j \leq m$  satisfying  $m^{2a+2}/n = o(1)$ ,  $m^{2a+2}/(n\bar{N}_2^2h^2) = o(1)$ ,  $h^4m^{2a+2c} \lesssim 1$  and  $hm^a \log n \lesssim 1$ ,*

$$(12) \quad \begin{aligned} \|\hat{\phi}_j - \phi_j\|_\infty = & O \left( \frac{j}{\sqrt{n}} (\sqrt{\ln n} + \ln j) \left\{ 1 + \frac{j^a}{\bar{N}_2} + \sqrt{\frac{j^{a-1}}{\bar{N}_2 h}} \left( 1 + \sqrt{\frac{j^a}{\bar{N}_2}} \right) \right\} \right. \\ & \left. + j^a \left| \frac{\ln n}{n} \right|^{1-\frac{1}{\alpha}} \left| j^{1/2} + \frac{\ln n}{\bar{N}_2 h} \right|^{1-\frac{1}{\alpha}} h^{-\frac{1}{\alpha}} + h^2 j^{c+1} \log j \right) a.s. \end{aligned}$$

The contributions of Theorem 4 are two-fold. First, our result is the first to establish uniform convergence for eigenfunctions with diverging indices, providing a useful tool for the theoretical analysis of models involving FPCA and inverse issues. Second, the double truncation technique we introduced reduces the truncation bias, making our results applicable to all sampling schemes. For a fixed  $j$ , Corollary 6 below discusses the uniform convergence rate under different ranges of  $\bar{N}_2$ . When  $\alpha > 5/2$ , the truncation bias in equation (12) is dominated by the first two terms for all scenarios of  $\bar{N}_2$ . If  $\bar{N}_2/(n/\log n)^{1/4} \rightarrow 0$ ,  $\|\hat{\phi}_j - \phi_j\|_\infty$  admits a typical one-dimensional kernel smoothing rate that differs only by a  $\log n$  factor, consistent with the result in Hall, Müller and Wang (2006) and Li and Hsing (2010). In summary, the introduction of the double truncation technique facilitates the attainment of uniform convergence across all sampling schemes within functional data analysis. This advancement prompts more in-depth research, particularly in the analysis of non-compact data exhibiting in-curve dependence.

**COROLLARY 6.** *Under the assumptions of Theorem 4. If  $j$  is fixed and  $\alpha > 5/2$ ,*

(1). *When  $\bar{N}_2/(n/\log n)^{1/4} \rightarrow 0$  and  $h \asymp (n\bar{N}_2/\log n)^{-1/5}$ ,*

$$\sup_{t \in [0,1]} |\hat{\phi}_j(t) - \phi_j(t)| = O \left( \left( \frac{\log n}{n\bar{N}_2} \right)^{\frac{2}{5}} \right) a.s.$$

(2). *When  $\bar{N}_2/(n/\log n)^{1/4} \rightarrow C_2 > 0$  and  $h \asymp (n/\log n)^{1/4}$*

$$\sup_{t \in [0,1]} |\hat{\phi}_j(t) - \phi_j(t)| = O \left( \sqrt{\frac{\log n}{n}} \right) a.s.$$

(3). *When  $\bar{N}_2/(n/\log n)^{1/4} \rightarrow \infty$ ,  $h = o(n/\log n)^{1/4}$  and  $\bar{N}_2 h \rightarrow \infty$ ,*

$$\sup_{t \in [0,1]} |\hat{\phi}_j(t) - \phi_j(t)| = O \left( \sqrt{\frac{\log n}{n}} \right) a.s.$$

The following corollary established the optimal uniform convergence rate for eigenfunctions with diverging indices under mild assumptions, a finding that is new even in the fully observed scenario. In contrast to the  $\mathcal{L}^2$  convergence, the maximum number of eigenfunctions that can be well-estimated under the  $\|\cdot\|_\infty$  norm is slightly smaller and depends on the moment characterized by assumption  $\alpha$ . If  $\alpha > 5/2$ , then  $m_{\max}$  can increase as the sample size  $n$  goes to infinity, and the phase transition point  $\bar{N}_2 \geq m_{\max}^a$  aligns with the  $\mathcal{L}^2$  case.

**COROLLARY 7.** *Under the assumptions of Theorem 4 and 5. Given  $\alpha > 5/2$ , denote  $m_{\max} = \min\{n^{\frac{1}{2\alpha}}, n^{\frac{\alpha-5/2}{2\alpha-\alpha-1}}\}$ . If  $\bar{N}_2 \geq m_{\max}^a$ ,  $h^4 m_{\max}^{2c} \leq n^{-1}$  and  $\bar{N}_2 h \geq m_{\max}$ , for all  $j \leq m_{\max}$ ,*

$$\sup_{t \in [0,1]} |\hat{\phi}_j(t) - \phi_j(t)| = O \left( \frac{j}{\sqrt{n}} (\sqrt{\ln n} + \ln j) \right) a.s.$$

**6. Numerical experiment.** In this section, we carry out a numerical evaluation of the convergence rates of eigenfunctions. The underlying trajectories are generated as  $X_i(t) = \sum_{j=1}^{50} \xi_{ij} \phi_j(t)$ , for  $i = 1, \dots, n$ . The principal component scores are independently generated following the distribution  $\xi_{ij} \sim \mathcal{N}(0, j^{-2})$  for all  $i$  and  $j$ . We define the eigenfunctions as  $\phi_1(t) \equiv 1$ , and for  $j \geq 1$ , as  $\phi_j(t) = \sqrt{2} \cos(j\pi t)$ . The actual observations are  $X_{ij} = X_i(t_{ij}) + \epsilon_{ij}$ , with noise  $\epsilon_{ij}$  following  $\mathcal{N}(0, 0.1^2)$ , and the time points  $t_{ij}$  sampled from a uniform distribution  $\text{Unif}[0, 1]$ , for  $j = 1, \dots, N$ . Each setting is repeated for 200 Mento-Carlo runs to mitigate the randomness that may occur in a single simulation.

When the phase transition occurs, our theory indicates a proportional relationship such that  $\log(\|\hat{\phi}_j - \phi_j\|^2) \propto 2 \log(j)$  for each fixed  $n$  and  $\log(\|\hat{\phi}_j - \phi_j\|^2) \propto -\log(n)$  for each fixed  $j$ . Figure 1 illustrates this phenomenon by showing that as  $N$  increases, the relationship between  $\log(\|\hat{\phi}_j - \phi_j\|^2)$  and  $j$  tends to be linear with a slope of 2, indicating that the phase transition might occur around  $N = 50$ . Figure 1 additionally offers a practical way for identifying the phase transition point of  $N$ . This is valuable in guiding both data collection and experimental design, contributing to more cost-effective data collecting strategies. Similarly, Figure 2 shows that as  $N$  increases, the relationship between  $\log(\|\hat{\phi}_j - \phi_j\|^2)$  and  $\log(n)$  also tends to follow a linear trend, but with a slope of  $-1$ .

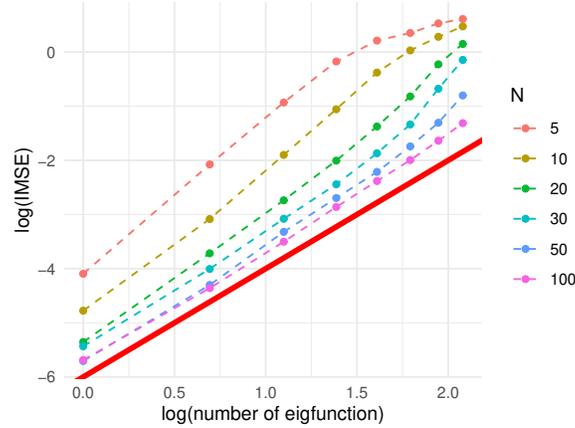


FIG 1. Plot of  $\log(\|\hat{\phi}_j - \phi_j\|^2)$  over the logarithm of number of eigenfunctions. The sample size is  $n = 240$ . The colored dashed lines correspond to different value of  $N$ . The solid red line represents the theoretical optimal value.

We compare the integrated mean square error  $\|\hat{\phi}_j - \phi_j\|^2$  for  $j = 1, 2, 3$ , between the pooling and pre-smoothing methods. The data are generated as before, using four different sample sizes ( $n = 100, 260, 240, 500$ ) and three different sampling schemes ( $N = 30, 50, 100$ ). The pre-smoothing method is implemented using the R package `fda` (Ramsay, Hooker and Graves, 2024), with each subject pre-smoothed by `smooth.basis()` employing 20 Fourier basis functions. The results summarized in Table 1 indicate that the pooling method consistently outperforms the pre-smoothing method in all settings when  $N = 30$ . In the ultra-dense cases ( $N = 50$  and  $N = 100$ ), the pooling method performs as well as the pre-smoothing method for larger sample sizes and shows superior performance for relatively small sample sizes. This suggests the use of pooling estimation in most cases, given its efficiency in both sparse and dense schemes.

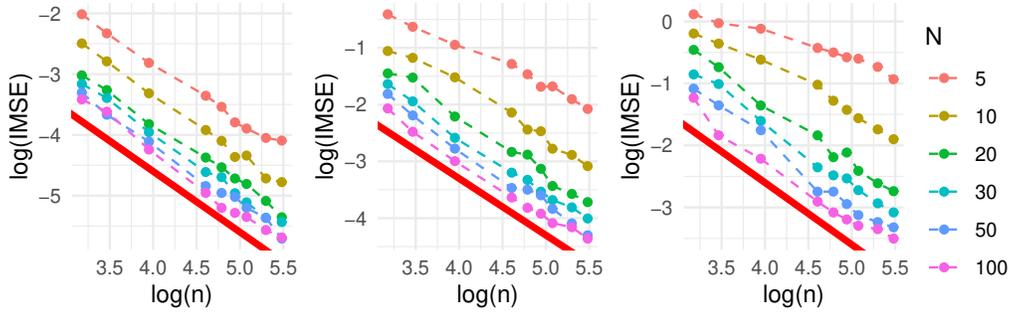


FIG 2. Plot of  $\log(\|\hat{\phi}_j - \phi_j\|^2)$  for the first (left), second (middle) and third (right) eigenfunctions over  $\log(n)$ . The colored dashed lines correspond to different value of  $N$ . The solid red line represents the theoretical optimal value.

TABLE 1

$\|\hat{\phi}_j - \phi_j\|^2$  for  $j = 1, 2, 3$  across different sample sizes and sampling schemes using pooling and pre-smoothing methods, respectively, where all values are multiplied by  $10^2$  for better visualization.

| Methods      | 1st   |       |       | 2nd   |       |       | 3rd   |       |       |       |
|--------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
|              | N=30  | N=50  | N=100 | N=30  | N=50  | N=100 | N=30  | N=50  | N=100 |       |
| Pooling      | n=100 | .9845 | .8551 | .7617 | 4.184 | 3.153 | 2.523 | 10.73 | 6.774 | 5.343 |
|              | n=160 | .6970 | .5228 | .4895 | 2.599 | 2.090 | 1.817 | 5.902 | 4.460 | 3.563 |
|              | n=240 | .4514 | .4738 | .3078 | 1.814 | 1.622 | 1.342 | 4.544 | 3.653 | 2.988 |
|              | n=500 | .2092 | .1938 | .1472 | 1.091 | .8426 | .7179 | 2.961 | 2.240 | 2.021 |
| Presmoothing | n=100 | 1.433 | .8591 | .7641 | 4.382 | 3.318 | 2.689 | 11.21 | 6.780 | 5.348 |
|              | n=160 | 1.230 | .5239 | .4896 | 3.143 | 2.270 | 1.981 | 5.986 | 4.481 | 3.566 |
|              | n=240 | .9109 | .4739 | .3092 | 2.308 | 1.792 | 1.504 | 4.666 | 3.663 | 2.989 |
|              | n=500 | .5403 | .1942 | .1473 | 2.086 | 1.005 | .8799 | 3.838 | 2.240 | 2.022 |

**7. Conclusion and discussion.** In this paper, we focus on the convergence rate of eigenfunctions with diverging indices for discretely observed functional data. We propose new techniques to handle the perturbation series and establish sharp bounds for eigenvalues and eigenfunctions across different convergence types. Additionally, we extend the partition “dense” and “sparse” defined for mean and covariance functions to principal components analysis. Another notable contribution of this paper is the double truncation technique for handling uniform convergence. Existing results on uniform convergence for covariance estimation require a strong moment condition on  $X(t)$  and are only applicable to sparse functional data where  $N_i = o_p(n^{1/4})$ . By employing the double truncation technique proposed in this paper, we establish an improved bound for the truncated bias, which ensures the uniform convergence of the covariance and eigenfunctions across all sampling schemes under mild conditions. These asymptotic properties play a direct role in various types of statistical inference involving functional data (Yao, Müller and Wang, 2005a; Li and Hsing, 2010).

Furthermore, the optimal rate achieved in this paper holds significant implications for downstream analysis. Since most functional regression models encounter inverse issues due to the infinite dimensionality of functional covariates, the convergence rates in this paper would help improve existing theoretical findings in downstream analyses from fully observed functional data to various sampling designs. Consider the functional linear model  $\mathbb{E}[Y_i | X_i] = \int \beta X_i$  with  $\beta = \sum_{k=1}^{\infty} k^{-b} \phi_k$ . Without a sharp bound for eigenfunctions, achieving optimal

convergence for standard and efficient plug-in estimators becomes challenging. Therefore, methods like approximated least squares and sample splitting, as discussed in [Zhou, Yao and Zhang \(2023\)](#), are necessary in the modeling phase. These complex methods require estimating principal components scores and do not efficiently utilize information gained from pooling. As a result, the phase transition for the functional linear model obtained by [Zhou, Yao and Zhang \(2023\)](#) is  $\max\{n^{\frac{2a+2}{a+2b}}, n^{\frac{a+b}{a+2b}}\}$ , which is significantly greater than  $1/2$ . In contrast, by using the new results in this paper, one can directly apply the plug-in estimator from [Hall and Horowitz \(2007\)](#) and achieve a phase transition of  $n^{\frac{1}{4} + \frac{3a-2b}{4(a+2b)}}$  for the functional linear model. Additionally, for complex regression models like the functional generalized linear model or functional Cox model, the methods developed in this paper could serve as a cornerstone for further exploration.

#### APPENDIX A: PROOF OF THEOREM 1

In this section, we outline the proof of Theorem 1 under the random design setting. The detailed proofs of the auxiliary results, as well as the case of fixed design, are provided in the Supplement.

Recall  $\hat{C}(s, t) = \hat{\beta}_0$  in (4) and  $\hat{\beta}_0$  has the analytical form

$$\hat{\beta}_0(s, t) = \frac{R_{00}(s, t)I_1(s, t) + R_{10}(s, t)I_2(s, t) + R_{01}(s, t)I_3(s, t)}{S_{00}(s, t)I_1(s, t) + S_{10}(s, t)I_2(s, t) + S_{01}(s, t)I_3(s, t)},$$

where

$$S_{pq}(s, t) = \sum_{i=1}^n v_i \sum_{l_1 \neq l_2}^{N_i} \frac{1}{h^2} \text{K} \left( \frac{t_{il_1} - s}{h} \right) \text{K} \left( \frac{t_{il_2} - t}{h} \right) \left( \frac{t_{il_1} - s}{h} \right)^p \left( \frac{t_{il_2} - t}{h} \right)^q;$$

$$R_{pq}(s, t) = \sum_{i=1}^n v_i \sum_{l_1 \neq l_2}^{N_i} \frac{1}{h^2} \text{K} \left( \frac{t_{il_1} - s}{h} \right) \text{K} \left( \frac{t_{il_2} - t}{h} \right) \left( \frac{t_{il_1} - s}{h} \right)^p \left( \frac{t_{il_2} - t}{h} \right)^q \delta_{il_1 l_2}$$

for  $p, q = 0, 1, 2$  and

$$I_1(s, t) = S_{02}(s, t)S_{20}(s, t) - S_{11}^2(s, t), \quad I_2(s, t) = S_{01}(s, t)S_{11}(s, t) - S_{02}(s, t)S_{10}(s, t)$$

$$I_3(s, t) = S_{10}(s, t)S_{11}(s, t) - S_{20}(s, t)S_{01}(s, t).$$

Similar calculations show that

$$\hat{\beta}_1(s, t) = \frac{1}{h} \frac{R_{00}(s, t)I_2(s, t) + R_{10}(s, t)J_1(s, t) + R_{01}(s, t)J_3(s, t)}{S_{00}(s, t)I_1(s, t) + S_{10}(s, t)I_2(s, t) + S_{01}(s, t)I_3(s, t)}$$

and

$$\hat{\beta}_2(s, t) = \frac{1}{h} \frac{R_{00}(s, t)I_3(s, t) + R_{10}(s, t)J_3(s, t) + R_{01}(s, t)J_2(s, t)}{S_{00}(s, t)I_1(s, t) + S_{10}(s, t)I_2(s, t) + S_{01}(s, t)I_3(s, t)}$$

where

$$J_1(s, t) = S_{00}(s, t)S_{02}(s, t) - S_{01}^2(s, t), \quad J_2(s, t) = S_{00}(s, t)S_{20}(s, t) - S_{10}^2(s, t)$$

$$J_3(s, t) = S_{10}(s, t)S_{01}(s, t) - S_{00}(s, t)S_{11}(s, t).$$

In the following, we omit the arguments  $(s, t)$  in the functions  $\hat{\beta}_1, \hat{\beta}_2, S_{pq}, R_{pq}, I_r$ , and  $J_r$  for  $p, q = 0, 1, 2$  and  $r = 1, 2, 3$ , whenever there is no ambiguity. Some calculations show that  $\hat{C}(s, t) = (R_{00} - h\hat{\beta}_1 S_{10} - h\hat{\beta}_2 S_{01})/S_{00}$ . We further define

$$\tilde{C}_0(s, t) = \left\{ R_{00} - C(s, t)S_{00} - h \frac{\partial C(s, t)}{\partial s}(s, t)S_{10} - h \frac{\partial C(s, t)}{\partial t}(s, t)S_{01} \right\} / f(s)f(t),$$

where  $f(s)$  is the density function of  $t_{ij}$ .

By the proof of Theorem 5.1.8 in [Hsing and Eubank \(2015\)](#), for  $j \in m$  and on the set  $\Omega_m(n, N, h) = \{\|\hat{C} - C\|_{\text{HS}} \leq \eta_m/2\}$ , the following expansion holds:

$$(13) \quad \begin{aligned} \hat{\phi}_j - \phi_j &= \sum_{k \neq j} \frac{f(\hat{C} - C)\phi_j\phi_k}{(\lambda_j - \lambda_k)}\phi_k + \sum_{k \neq j} \frac{f(\hat{C} - C)(\hat{\phi}_j - \phi_j)\phi_k}{(\lambda_j - \lambda_k)}\phi_k \\ &+ \sum_{k \neq j} \sum_{s=1}^{\infty} \frac{(\lambda_j - \hat{\lambda}_j)^s}{(\lambda_j - \lambda_k)^{s+1}} \left\{ \int (\hat{C} - C)\hat{\phi}_j\phi_k \right\} \phi_k + \left\{ \int (\hat{\phi}_j - \phi_j)\phi_j \right\} \phi_j. \end{aligned}$$

Such kind of expansion can also be found in [Hall and Hosseini-Nasab \(2006\)](#) and [Li and Hsing \(2010\)](#). Below, when we say that a bound is valid when  $\Omega_m(n, N, h)$  holds, this should be interpreted as stating that the bound is valid for all realizations for which  $\|\hat{C} - C\|_{\text{HS}} \leq \eta_m/2$  ([Hall and Horowitz, 2007](#)). Under assumptions  $m^{2a+2}/n \rightarrow 0$ ,  $m^{2a+2}/(n\bar{N}_2^2 h^2) \rightarrow 0$  and  $h^2 \max\{m^{a+c}, m^{4a} \log n\} \lesssim 1$ , we have  $\mathbb{P}(\Omega_m(n, N, h)) \rightarrow 1$ . Since  $\lim_{D \rightarrow \infty} \limsup_{n \rightarrow \infty} P(\|\hat{\phi}_j - \phi_j\| > D\tau_n) = 0$  implies  $\|\hat{\phi}_j - \phi_j\| = O_P(\tau_n)$  for any positive sequence  $\tau_n$ , the results in  $O_P$  form that we aim to prove pertain only to the probabilities of differences. Therefore, it suffices to work with bounds established under the assumption that  $\mathbb{P}(\Omega_m) \rightarrow 1$ , as discussed in Section 5.1 of [Hall and Horowitz \(2007\)](#).

We first show that  $\mathbb{E}(\|\hat{\phi}_j - \phi_j\|^2)$  is dominated by the  $\mathcal{L}^2$  norm of the first term in the right hand side of equation (13). By Bessel's inequality, we see that

$$(14) \quad \mathbb{E} \left\| \sum_{k \neq j} \frac{f(\hat{C} - C)(\hat{\phi}_j - \phi_j)\phi_k}{(\lambda_j - \lambda_k)} \phi_k \right\|^2 \leq \mathbb{E} \frac{\|\hat{C} - C\|_{\text{HS}}^2 \|\hat{\phi}_j - \phi_j\|^2}{(2\eta_j)^2} < \frac{1}{16} \mathbb{E} \|\hat{\phi}_j - \phi_j\|^2,$$

where the last equality comes from the fact  $\eta_j^{-1} \|\hat{C} - C\| < 1/2$  on  $\Omega_m(n, N, h)$ . Similarly,

$$(15) \quad \begin{aligned} &\mathbb{E} \left\| \sum_{k \neq j} \sum_{s=1}^{\infty} \frac{(\lambda_j - \hat{\lambda}_j)^s}{(\lambda_j - \lambda_k)^{s+1}} \left\{ \int (\hat{C} - C)\hat{\phi}_j\phi_k \right\} \phi_k \right\|^2 \\ &= \mathbb{E} \sum_{k \neq j} \frac{(\lambda_j - \hat{\lambda}_j)^2}{(\lambda_j - \lambda_k)^2 (\hat{\lambda}_j - \lambda_k)^2} \left\{ \int (\hat{C} - C)\hat{\phi}_j\phi_k \right\}^2 \\ &\leq 2\mathbb{E} \frac{\|\hat{C} - C\|_{\text{HS}}^2}{(2\eta_j - \|\hat{C} - C\|_{\text{HS}})^2} \left[ \sum_{k \neq j} \frac{\left\{ \int (\hat{C} - C)\phi_j\phi_k \right\}^2}{(\lambda_j - \lambda_k)^2} + \sum_{k \neq j} \frac{\left\{ \int (\hat{C} - C)(\hat{\phi}_j - \phi_j)\phi_k \right\}^2}{(\lambda_j - \lambda_k)^2} \right] \\ &\leq \frac{8}{9} \mathbb{E} \left[ \frac{\|\hat{C} - C\|_{\text{HS}}^2}{\eta_j^2} \sum_{k \neq j} \frac{\left\{ \int (\hat{C} - C)\phi_j\phi_k \right\}^2}{(\lambda_j - \lambda_k)^2} + \frac{\|\hat{C} - C\|_{\text{HS}}^4}{\eta_j^4} \|\hat{\phi}_j - \phi_j\|^2 \right] \\ &\leq \frac{2}{9} \mathbb{E} \sum_{k \neq j} \frac{\left\{ \int (\hat{C} - C)\phi_j\phi_k \right\}^2}{(\lambda_j - \lambda_k)^2} + \frac{1}{18} \mathbb{E} \|\hat{\phi}_j - \phi_j\|^2. \end{aligned}$$

Combine (13) to (15) and the fact  $\|\{ \int (\hat{\phi}_j - \phi_j)\phi_j \} \phi_j\| = 1/2 \|\hat{\phi}_j - \phi_j\|^2$ ,  $\mathbb{E}(\|\hat{\phi}_j - \phi_j\|^2)$  is dominated by the first term in the right hand side of equation (13). Thus,

$$(16) \quad \mathbb{E}(\|\hat{\phi}_j - \phi_j\|^2) \lesssim \sum_{k \neq j} \frac{\mathbb{E} \left\{ \left( \int \tilde{C}_0 \phi_j \phi_k \right)^2 \right\}}{(\lambda_j - \lambda_k)^2} + \sum_{k \neq j} \frac{\mathbb{E} \left[ \left\{ \int (\hat{C} - C - \tilde{C}_0) \phi_j \phi_k \right\}^2 \right]}{(\lambda_j - \lambda_k)^2}.$$

We focus on the first term on the right-hand side of equation (16). As discussed in Section 2, the summation  $\sum_{k \neq j} \mathbb{E}\{(\int \tilde{C}_0 \phi_j \phi_k)^2\}/(\lambda_j - \lambda_k)^2$  should be evaluated separately over the sets  $S_1 = \{k \leq 2j, k \neq j\}$  and  $S_2 = \{k > 2j\}$ . The following lemma provides sharp bounds for  $\mathbb{E}\{(\int \tilde{C}_0 \phi_j \phi_k)^2\}$ , enabling precise estimates for the summation over the sets  $S_1$  and  $S_2$ .

LEMMA 1. *Under assumptions 1 to 4, 6,  $h^4 j^{2a+2c} \lesssim 1$  and  $h j^a \log n \lesssim 1$ , the following holds for all  $1 \leq k \leq 2j$ :*

$$\mathbb{E} \left[ \left\{ \iint \tilde{C}_0(s, t) \phi_j(s) \phi_k(t) ds dt \right\}^2 \right] \lesssim \frac{1}{n} \left( j^{-a} k^{-a} + \frac{j^{-a} + k^{-a}}{\bar{N}_2} + \frac{1}{\bar{N}_2^2} \right) + h^4 k^{2c-2a}.$$

Additionally, for  $k > 2j$ ,

$$\begin{aligned} & \sum_{k=2j+1}^{\infty} \mathbb{E} \left[ \left\{ \iint \tilde{C}_0(s, t) \phi_j(s) \phi_k(t) ds dt \right\}^2 \right] \\ & \lesssim \frac{1}{n} \left( j^{1-2a} + \frac{h^{-1} j^{-a} + j^{1-a}}{\bar{N}_2^2} + \frac{1}{h \bar{N}_2^2} \right) + h^4 j^{1+2c-2a}. \end{aligned}$$

Then, the first term in the right hand side of (16) is bounded by

$$\begin{aligned} & \sum_{k \neq j} \frac{\mathbb{E} \left\{ \left( \int \tilde{C}_0 \phi_j \phi_k \right)^2 \right\}}{(\lambda_j - \lambda_k)^2} = \sum_{\substack{k \neq j \\ k \leq 2j}} \frac{\mathbb{E} \left\{ \left( \int \tilde{C}_0 \phi_j \phi_k \right)^2 \right\}}{(\lambda_j - \lambda_k)^2} + \sum_{k > 2j} \frac{\mathbb{E} \left\{ \left( \int \tilde{C}_0 \phi_j \phi_k \right)^2 \right\}}{(\lambda_j - \lambda_k)^2} \\ & \lesssim h^4 j^{2c+2} + \frac{j^2}{n} \left\{ 1 + \frac{j^a}{\bar{N}_2} + \frac{j^{2a}}{\bar{N}_2^2} \right\} + h^4 j^{2c+1} + \frac{1}{n} \left( j + \frac{j^a}{\bar{N}_2 h} + \frac{j^{a+1}}{\bar{N}_2} + \frac{j^{2a}}{\bar{N}_2^2 h} \right) \\ & \lesssim \frac{j^2}{n} \left\{ 1 + \frac{j^a}{\bar{N}_2} + \frac{j^{2a}}{\bar{N}_2^2} \right\} + \frac{j^a}{nh} \left( \frac{1}{\bar{N}_2} + \frac{j^a}{\bar{N}_2^2} \right) + h^4 j^{2c+2}, \end{aligned}$$

where the second inequality comes from Lemma 7 in Dou, Pollard and Zhou (2012) and Lemma 1.

By Bessel equality, the second term in the right hand side of equation (16) is bounded by

$$\sum_{k \neq j} \frac{\mathbb{E} \left\{ \int (\hat{C} - C - \tilde{C}_0) \phi_j \phi_k \right\}^2}{(\lambda_j - \lambda_k)^2} \leq \frac{\left\| \hat{C} - C - \tilde{C}_0 \right\|_{\text{HS}}^2}{\eta_j^2} \lesssim j^{2a+2} \left\| \hat{C} - C - \tilde{C}_0 \right\|_{\text{HS}}^2.$$

Note that

$$\begin{aligned} & \hat{C}(s, t) - C(s, t) - \tilde{C}_0(s, t) \\ (17) \quad & = \left[ -h \left\{ \hat{\beta}_1 - \frac{\partial C(s, t)}{\partial s} \right\} S_{10} - h \left\{ \hat{\beta}_2 - \frac{\partial C(s, t)}{\partial t} \right\} S_{01} \right] / S_{00} \\ & \quad + \left\{ R_{00} - C(s, t) S_{00} - h \frac{\partial C(s, t)}{\partial s} S_{10} - h \frac{\partial C(s, t)}{\partial t} S_{01} \right\} \left\{ \frac{1}{S_{00}} - \frac{1}{f(s)f(t)} \right\}. \end{aligned}$$

For the first part in the right hand side of (17), some calculations show that

$$\begin{aligned} & h \left\{ \hat{\beta}_1 - \frac{\partial C(s, t)}{\partial s} \right\} S_{10} + h \left\{ \hat{\beta}_2 - \frac{\partial C(s, t)}{\partial t} \right\} S_{01} \\ (18) \quad & = \frac{S_{10} I_2 + S_{01} I_3}{S_{00} I_1 + S_{10} I_2 + S_{01} I_3} \left\{ R_{00} - C(s, t) S_{00} - h \frac{\partial C(s, t)}{\partial s} S_{10} - h \frac{\partial C(s, t)}{\partial t} S_{01} \right\} \\ & \quad + \frac{S_{10} J_1 + S_{01} J_3}{S_{00} I_1 + S_{10} I_2 + S_{01} I_3} \left\{ R_{10} - C(s, t) S_{10} - h \frac{\partial C(s, t)}{\partial s} S_{20} - h \frac{\partial C(s, t)}{\partial t} S_{11} \right\} \\ & \quad + \frac{S_{10} J_3 + S_{01} J_2}{S_{00} I_1 + S_{10} I_2 + S_{01} I_3} \left\{ R_{01} - C(s, t) S_{01} - h \frac{\partial C(s, t)}{\partial s} S_{11} - h \frac{\partial C(s, t)}{\partial t} S_{02} \right\}. \end{aligned}$$

To bound each term in the right hand side of equation (18), we need the following proposition.

PROPOSITION 1. *Under Assumption 2 and 6,*

(a)  $\inf_{s,t} S_{00}(s,t)I_1(s,t) + S_{10}(s,t)I_2(s,t) + S_{01}(s,t)I_3(s,t)$  and  $\inf_{s,t} S_{00}(s,t)$  are bounded away from zero almost surely.

(b)

$$\|S_{00}(s,t) - f(s)f(t)\|_{\text{HS}}^2 = O_P \left( h^2 + \frac{1}{n} \left\{ 1 + \frac{1}{N_2^2 h^2} \right\} \right).$$

(c) For  $p, q = 0, 1, 2$ ,  $\|S_{pq}\|_{\infty} = O(1)$  a.s.

(d) For  $p, q = 0, 1$ ,  $p + q = 1$

$$\|S_{pq}(s,t)\|_{\text{HS}}^2 = O_P \left( h + \frac{\log n}{n} \left\{ 1 + \frac{1}{N_2^2 h^2} \right\} \right).$$

(e) In addition, if Assumption 5 holds with  $\alpha > 3$ , for  $p, q = 0, 1$

$$\begin{aligned} & \sup_{s,t} \left| R_{pq}(s,t) - C(s,t)S_{pq}(s,t) - h \frac{\partial C(s,t)}{\partial s} S_{p+1,q}(s,t) - h \frac{\partial C(s,t)}{\partial t} S_{p,q+1}(s,t) \right| \\ &= O \left( h^2 + \sqrt{\frac{\log n}{n} \left\{ 1 + \frac{1}{N_2^2 h^2} \right\}} \right) \text{ a.s.} \end{aligned}$$

By equation (17) and (18), the HS-norm of  $\hat{C}(s,t) - C(s,t) - \tilde{C}_0$  can be bounded by

$$\begin{aligned} & \left\| \hat{C}(s,t) - C(s,t) - \tilde{C}_0 \right\|_{\text{HS}}^2 \\ & \lesssim \sup_{s,t} \frac{\|S_{10}\|_{\text{HS}}^2 I_2^2 + \|S_{01}\|_{\text{HS}}^2 I_3^2}{S_{00}^2 (S_{00} I_1 + S_{10} I_2 + S_{01} I_3)^2} \\ & \quad \times \left\| R_{00} - C(s,t)S_{00} - h \frac{\partial C(s,t)}{\partial s} S_{10} - h \frac{\partial C(s,t)}{\partial t} S_{01} \right\|_{\infty}^2 \\ & + \sup_{s,t} \frac{\|S_{10}\|_{\text{HS}}^2 J_1^2 + \|S_{01}\|_{\text{HS}}^2 J_3^2}{S_{00}^2 (S_{00} I_1 + S_{10} I_2 + S_{01} I_3)^2} \\ & \quad \times \left\| R_{10} - C(s,t)S_{10} - h \frac{\partial C(s,t)}{\partial s} S_{20} - h \frac{\partial C(s,t)}{\partial t} S_{11} \right\|_{\infty}^2 \\ & + \sup_{s,t} \frac{\|S_{10}\|_{\text{HS}}^2 J_3^2 + \|S_{01}\|_{\text{HS}}^2 J_2^2}{S_{00}^2 (S_{00} I_1 + S_{10} I_2 + S_{01} I_3)^2} \\ & \quad \times \left\| R_{01} - C(s,t)S_{01} - h \frac{\partial C(s,t)}{\partial s} S_{11} - h \frac{\partial C(s,t)}{\partial t} S_{02} \right\|_{\infty}^2 \\ & + \sup_{s,t} \frac{\left| R_{00} - C(s,t)S_{00} - h \frac{\partial C(s,t)}{\partial s} S_{10} - h \frac{\partial C(s,t)}{\partial t} S_{01} \right|^2}{S_{00}^2 f(s)^2 f(t)^2} \|S_{00} - f(s)f(t)\|_{\text{HS}}^2. \\ & = O_P \left( h^5 + \frac{h}{n} \left\{ 1 + \frac{\log n}{N_2^2 h^2} \right\} \right). \end{aligned} \tag{19}$$

Since under assumption  $h j^{2a} \log n = o(1)$ ,  $j^{2a+2} \|\hat{C} - C - \tilde{C}_0\|_{\text{HS}}^2$  is dominated by the first term of equation (16), this completes the proof.

APPENDIX B: DOUBLE TRUNCATION TECHNIQUE IN UNIFORM CONVERGENCE.

In obtaining the uniform convergence of the mean and covariance functions for functional data, a key quantity is the rate of  $\sup_{s,t \in [0,1]} |R_{pq}(s,t) - \mathbb{E}R_{pq}(s,t)|$ . To handle the uniform rate on a continuous interval like  $[0, 1]$ , a well-developed technique (Li and Hsing, 2010; Zhang and Wang, 2016) involves considering a mesh grid  $\chi(\rho) = \{n^{-\rho}(i, j) \mid i, j \in \mathbb{Z} \cap (0, n^\rho)\}$  on  $[0, 1]^2$  with a gap of  $n^{-\gamma}$ . The rate can then be decomposed as

$$(20) \quad \sup_{s,t \in [0,1]} |R_{00}(s,t) - \mathbb{E}R_{00}(s,t)| \leq \sup_{(s,t) \in \chi(\rho)} |R_{00}(s,t) - \mathbb{E}R_{00}(s,t)| + D,$$

where the second term can be proven to be negligible due to the smoothness of the kernel function.

Rewrite  $R_{00} - \mathbb{E}R_{00} = \sum_{i=1}^n V_i$ , where

$$V_i = \frac{1}{nN_i(N_i - 1)} \sum_{1 \leq l_1 \neq l_2 \leq N_i} \frac{1}{h^2} \mathbf{K}\left(\frac{t_{il_1} - s}{h}\right) \mathbf{K}\left(\frac{t_{il_2} - t}{h}\right) \delta_{il_1 l_2}$$

are zero-mean independent random variables. A common technique to establish the rate for the first term on the right-hand side of (20) is to use the inequality

$$\mathbb{P}\left(\sup_{(s,t) \in \chi(\rho)} |R_{00}(s,t) - \mathbb{E}R_{00}(s,t)| \geq Ma_n\right) \lesssim 2n^{2\rho} \mathbb{P}\left(\left|\sum_{i=1}^n V_i\right| \geq Ma_n\right),$$

where  $a_n$  is the desired rate and  $M$  is a larger constant. The probability  $\mathbb{P}(|\sum_{i=1}^n V_i| \geq Ma_n)$  can then be bounded using Bernstein's inequality. Finally, Borel–Cantelli's Lemma is applied to show that the event

$$\left\{ \sup_{(s,t) \in \chi(\rho)} |R_{00}(s,t) - \mathbb{E}R_{00}(s,t)| \geq Ma_n \right\}$$

occurs almost surely.

However, due to the unboundedness of functional data,  $\delta_{il_1 l_2} = X_{il_1} X_{il_2}$  is unbounded. To apply Bernstein's inequality, it is necessary to truncate  $\delta_{il_1 l_2}$  with a sequence  $A_n \rightarrow \infty$  as  $n \rightarrow \infty$ . Specifically, we have

$$(21) \quad \sup_{(s,t) \in \chi(\rho)} |R_{00}(s,t) - \mathbb{E}R_{00}(s,t)| \leq \sup_{(s,t) \in \chi(\rho)} |\tilde{R}_{00}(s,t) - \mathbb{E}\tilde{R}_{00}(s,t)| + E,$$

where  $\tilde{R}_{00}(s,t) = \sum_{i=1}^n V_i \mathbf{1}_{(|\delta_{il_1 l_2}| \leq A_n)}$ , and  $E$  represents the truncation bias. The uniform convergence of  $\sup_{(s,t) \in \chi(\rho)} |R_{00}(s,t) - \mathbb{E}R_{00}(s,t)|$  can be achieved by choosing an optimal  $A_n$  such that the two terms on the right-hand side of (21) are balanced. Once the optimal  $A_n$  is determined, it becomes essential to impose additional moment conditions on both  $X_i$  and  $\varepsilon_i$  to ensure that the truncation bias  $E$  is negligible.

To apply Bernstein's inequality to  $\tilde{R}_{00}(s,t)$ , it is necessary to bound  $\tilde{V}_i = V_i \mathbf{1}_{(|\delta_{il_1 l_2}| \leq A_n)}$  and its second-order moment. A straightforward bound for  $\tilde{V}_i$  is given by

$$(22) \quad \tilde{V}_i = \frac{1}{nN_i(N_i - 1)} \sum_{1 \leq l_1 \neq l_2 \leq N_i} \frac{1}{h^2} \mathbf{K}\left(\frac{t_{il_1} - s}{h}\right) \mathbf{K}\left(\frac{t_{il_2} - t}{h}\right) \delta_{il_1 l_2} \mathbf{1}_{(|\delta_{il_1 l_2}| \leq A_n)} \leq \frac{A_n}{h^2 n}.$$

However, using this rough bound causes the first term on the right-hand side of (21) to become too large, requiring  $A_n$  to be relatively small. As a consequence, relatively small  $A_n$  requires a strong moment assumptions on  $X_i$  and  $\varepsilon_i$  to make the truncation bias negligible. By applying the bound in (22), the current state-of-the-art results (Zhang and Wang, 2016) require the  $6/(1 - \tau)$ th moments of  $X_i$  and  $\varepsilon_i$  to be finite, where  $\tau$  is the sampling rate, and  $\bar{N}_2 \asymp n^{\tau/4}$  for  $\tau \in [0, 1)$ . However, as  $\bar{N}_2$  approaches  $n^{1/4}$ , the value of  $6/(1 - \tau)$  tends to infinity. Furthermore, in dense and ultra-dense sampling schemes where  $\bar{N}_2 \gtrsim n^{1/4}$ , the discrepancy between the truncated and the original estimators becomes dominant, preventing the achievement of optimal convergence rates with the current methods.

To address this issue and establish a unified theory for determining the uniform convergence rate of the covariance function, we propose a double truncation technique to obtain a sharper bound for  $\tilde{V}_i$ . Note that

$$\tilde{V}_i \leq A_n \frac{1}{n} \frac{N_i}{N_i - 1} J_i(s) J_i(t),$$

where  $J_i(s) := h^{-1} \sum_{l=1}^{N_i} K((t_{il} - s)/h)$ . For any fixed  $M > 5$ , Using Bernstein's inequality again gives

$$(23) \quad \mathbb{P}(J_i(s) > M) \leq \exp\left(-\frac{MN_i h}{3}\right).$$

Define

$$\tilde{\tilde{V}}_i = \tilde{V}_i \mathbf{1}_{(|\tilde{V}_i(s,t)| \leq A_n \frac{1}{n} \frac{N_i M^2}{N_i - 1})} \text{ and } \tilde{R}_{00}(s,t) = \sum_{i=1}^n \tilde{\tilde{V}}_i$$

then

$$(24) \quad \sup_{(s,t) \in \chi(\rho)} |\tilde{R}_{00}(s,t) - \mathbb{E}\tilde{R}_{00}(s,t)| \leq \sup_{(s,t) \in \chi(\rho)} |\tilde{\tilde{R}}_{00}(s,t) - \mathbb{E}\tilde{\tilde{R}}_{00}(s,t)| + F.$$

The first term on the right-hand side of (24) can be bounded using Bernstein's inequality with the bound  $\tilde{\tilde{V}}_i \lesssim A_n M^2/n$ . The bias term  $F$  caused by the second truncation can also be bounded using (23). The final convergence rate is obtained by selecting optimal values for  $M$  and  $A_n$  to balance  $\sup_{(s,t) \in \chi(\rho)} |\tilde{\tilde{R}}_{00}(s,t) - \mathbb{E}\tilde{\tilde{R}}_{00}(s,t)|$ , the first truncation error  $E$ , and the second truncation bias  $F$ .

The detailed proof of Theorem 3 can be found in the Supplement, where the optimal convergence rate is established for both sparse and dense sampling schemes under mild moment assumptions. In summary, using the double truncation technique, we decompose the previous dominating term  $\sup_{(s,t) \in \chi(\rho)} |\tilde{R}_{00}(s,t) - \mathbb{E}\tilde{R}_{00}(s,t)|$  into two components:  $\sup_{(s,t) \in \chi(\rho)} |\tilde{\tilde{R}}_{00}(s,t) - \mathbb{E}\tilde{\tilde{R}}_{00}(s,t)|$  and  $F$ . By applying a sharp bound to  $F$ , the term  $\sup_{(s,t) \in \chi(\rho)} |\tilde{R}_{00}(s,t) - \mathbb{E}\tilde{R}_{00}(s,t)|$  becomes significantly smaller compared to using the bound in (22) directly.

**Funding.** This research is supported in part by the National Key R&D Program of China Grant 2022YFA1003801, the National Natural Science Foundation of China Grants 12292981 and 12288101, the New Cornerstone Science Foundation through the XPLOER PRIZE, the LMAM, the Fundamental Research Funds for the Central Universities, Peking University (LMEQF).

## SUPPLEMENTARY MATERIAL

The Supplement contains detailed proofs of the theorems presented in Sections 4 and 5, along with auxiliary results.

## REFERENCES

- ANDERSON, G. W., GUIONNET, A. and ZEITOUNI, O. (2010). *An Introduction to Random Matrices* **118**. Cambridge university press.
- BAI, Z. and SILVERSTEIN, J. W. (2010). *Spectral Analysis of Large Dimensional Random Matrices* **20**. Springer.
- BAI, Z., JIANG, D., YAO, J.-F. and ZHENG, S. (2009). Corrections to LRT on large-dimensional covariance matrix by RMT. *Ann. Statist.* **37** 3822–3840.
- BAO, Z., DING, X., WANG, J. and WANG, K. (2022). Statistical inference for principal components of spiked covariance matrices. *Ann. Statist.* **50** 1144–1169.
- BIANCHI, P., DEBBAH, M., MAÍDA, M. and NAJIM, J. (2011). Performance of statistical tests for single-source detection using random matrix theory. *IEEE Trans. Inf. Theory.* **57** 2400–2419.

- BICKEL, P. J. and ROSENBLATT, M. (1973). On some global measures of the deviations of density function estimates. *Ann. Statist.* 1071–1095.
- BOSQ, D. (2000). *Linear Processes in Function Spaces: Theory and Applications* **149**. Springer Science & Business Media.
- CAI, T. T. and HALL, P. (2006). Prediction in functional linear regression. *Ann. Statist.* **34** 2159–2179.
- CAI, T. and YUAN, M. (2010). Nonparametric covariance function estimation for functional and longitudinal data. *Technical report*.
- CAI, T. T. and YUAN, M. (2011). Optimal estimation of the mean function based on discretely sampled functional data: Phase transition. *Ann. Statist.* **39** 2330–2355.
- CHEN, S. X. and QIN, Y.-L. (2010). A two-sample test for high-dimensional data with applications to gene-set testing. *Ann. Statist.* **38** 808–835.
- DING, X. and YANG, F. (2021). Spiked separable covariance matrices and principal components. *Ann. Statist.* **49** 1113–1138.
- DOU, W. W., POLLARD, D. and ZHOU, H. H. (2012). Estimation in functional regression for general exponential families. *Ann. Statist.* **40** 2421–2451.
- EL KAROUI, N. (2007). Tracy–Widom limit for the largest eigenvalue of a large class of complex sample covariance matrices. *Ann. Probab.* **35** 663–714.
- FERRATY, F. and VIEU, P. (2006). *Nonparametric Functional Data Analysis: Theory and Practice*. Springer Science & Business Media.
- HALL, P. and HOROWITZ, J. L. (2007). Methodology and convergence rates for functional linear regression. *Ann. Statist.* **35** 70–91.
- HALL, P. and HOSSEINI-NASAB, M. (2006). On properties of functional principal components analysis. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **68** 109–126.
- HALL, P., MÜLLER, H.-G. and WANG, J.-L. (2006). Properties of principal component methods for functional and longitudinal data analysis. *Ann. Statist.* 1493–1517.
- HÄRDLE, W. (1989). Asymptotic maximal deviation of M-smoothers. *J. Multivar. Anal.* **29** 163–179.
- HÄRDLE, W., JANSSEN, P. and SERFLING, R. (1988). Strong uniform consistency rates for estimators of conditional functionals. *Ann. Statist.* 1428–1449.
- HORVÁTH, L. and KOKOSZKA, P. (2012). *Inference for Functional Data with Applications* **200**. Springer Science & Business Media.
- HSING, T. and EUBANK, R. (2015). *Theoretical Foundations of Functional Data Analysis, with an Introduction to Linear Operators* **997**. John Wiley & Sons.
- INDRITZ, J. (1963). *Methods in Analysis*. Macmillan.
- JOHNSTONE, I. M. (2001). On the distribution of the largest eigenvalue in principal components analysis. *Ann. Statist.* **29** 295–327.
- LI, Y. and HSING, T. (2010). Uniform convergence rates for nonparametric regression and principal component analysis in functional/longitudinal data. *Ann. Statist.* **38** 3321–3351.
- MÜLLER, H.-G. and STADTMÜLLER, U. (2005). Generalized functional linear models. *Ann. Statist.* **33** 774–805.
- NADLER, B., PENNA, F. and GARELLO, R. (2011). Performance of eigenvalue-based signal detectors with known and unknown noise level. In *2011 IEEE Int. Conf. Commun.* 1–5. IEEE.
- ONATSKI, A. (2009). Testing hypotheses about the number of factors in large factor models. *Econometrica* **77** 1447–1479.
- PASTUR, L. A. and SHCHERBINA, M. (2011). *Eigenvalue Distribution of Large Random Matrices* **171**. American Mathematical Soc.
- PAUL, D. (2007). Asymptotics of sample eigenstructure for a large dimensional spiked covariance model. *Stat. Sinica*. 1617–1642.
- PAUL, D. and AUE, A. (2014). Random matrix theory in statistics: A review. *J. Stat. Plan. Inference.* **150** 1–29.
- PAUL, D. and PENG, J. (2009). Consistency of restricted maximum likelihood estimators of principal components. *Ann. Statist.* **37** 1229–1271.
- QU, S., WANG, J.-L. and WANG, X. (2016). Optimal estimation for the functional cox model. *Ann. Statist.* **44** 1708–1738.
- RAMSAY, J., HOOKER, G. and GRAVES, S. (2024). Package ‘fda’.
- RAMSAY, J. and SILVERMAN, B. (2006). *Functional Data Analysis*. Springer Science & Business Media.
- RICE, J. A. and WU, C. O. (2001). Nonparametric mixed effects models for unequally sampled noisy curves. *Biometrics* **57** 253–259.
- SHAO, L., LIN, Z. and YAO, F. (2022). Intrinsic Riemannian functional data analysis for sparse longitudinal observations. *Ann. Statist.* **50** 1696 – 1721.
- WAHL, M. (2022). Lower bounds for invariant statistical models with applications to principal component analysis. In *Ann. Inst. Henri Poincaré Probab. Statist.* **58** 1565–1589. Institut Henri Poincaré.

- YAO, F. and LEE, T. C. (2006). Penalized spline models for functional principal component analysis. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **68** 3–25.
- YAO, F., MÜLLER, H.-G. and WANG, J.-L. (2005a). Functional data analysis for sparse longitudinal data. *J. Am. Stat. Assoc.* **100** 577–590.
- YAO, F., MÜLLER, H.-G. and WANG, J.-L. (2005b). Functional linear regression analysis for longitudinal data. *Ann. Statist.* 2873–2903.
- YUAN, M. and CAI, T. T. (2010). A reproducing kernel Hilbert space approach to functional linear regression. *Ann. Statist.* **38** 3412–3444.
- ZHANG, J.-T. and CHEN, J. (2007). Statistical inferences for functional data. *Ann. Statist.* **35** 1052–1079.
- ZHANG, X. and WANG, J.-L. (2016). From sparse to dense functional data and beyond. *Ann. Statist.* **44** 2281–2321.
- ZHENG, S. (2012). Central limit theorems for linear spectral statistics of large dimensional  $F$ -matrices. In *Ann. Inst. Henri Poincaré Probab. Statist* **48** 444–476.
- ZHOU, H., YAO, F. and ZHANG, H. (2023). Functional linear regression for discretely observed data: from ideal to reality. *Biometrika* **110** 381–393.